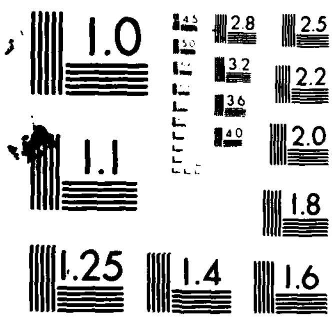
AD-A111 846 ADVISORY GROUP FOR AEROSPACE RESEARCH AND DEVELOPMENT--ETC F/G 5/2
WHAT SHOULD USERS EXPECT FROM INFORMATION STORAGE AND RETRIEVAL--ETC(U)
DEC 81
UNCLASSIFIED AGARD-CP-304 NL

1 of 2
AD
A111846



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

(1)

AGARD-CP-304

AGARD-CP-304

ADA111846

AGARD

ADVISORY GROUP FOR AEROSPACE RESEARCH & DEVELOPMENT

7 RUE ANCELLE 92200 NEUILLY SUR SEINE FRANCE

AGARD CONFERENCE PROCEEDINGS No. 304

What should Users Expect from Information Storage and Retrieval Systems of the 1980's?

NORTH ATLANTIC TREATY ORGANIZATION

DISTRIBUTION AND AVAILABILITY
ON BACK COVER

82 03 05 062

AGARD-CP-304

NORTH ATLANTIC TREATY ORGANIZATION

ADVISORY GROUP FOR AEROSPACE RESEARCH AND DEVELOPMENT

(ORGANISATION DU TRAITE DE L'ATLANTIQUE NORD)

AGARD Conference Proceedings No.304

WHAT SHOULD USERS EXPECT FROM INFORMATION STORAGE AND RETRIEVAL SYSTEMS OF THE 1980's?

Copies of papers presented at the Technical Information Panel Specialists' Meeting held in Munich, Federal Republic of Germany, 9–10 September 1981.

THE MISSION OF AGARD

The mission of AGARD is to bring together the leading personalities of the NATO nations in the fields of science and technology relating to aerospace for the following purposes:

- Exchanging of scientific and technical information;
- Continuously stimulating advances in the aerospace sciences relevant to strengthening the common defence posture;
- Improving the co-operation among member nations in aerospace research and development;
- Providing scientific and technical advice and assistance to the North Atlantic Military Committee in the field of aerospace research and development;
- Rendering scientific and technical assistance, as requested, to other NATO bodies and to member nations in connection with research and development problems in the aerospace field;
- Providing assistance to member nations for the purpose of increasing their scientific and technical potential;
- Recommending effective ways for the member nations to use their research and development capabilities for the common benefit of the NATO community.

The highest authority within AGARD is the National Delegates Board consisting of officially appointed senior representatives from each member nation. The mission of AGARD is carried out through the Panels which are composed of experts appointed by the National Delegates, the Consultant and Exchange Programme and the Aerospace Applications Studies Programme. The results of AGARD work are reported to the member nations and the NATO Authorities through the AGARD series of publications of which this is one.

Participation in AGARD activities is by invitation only and is normally limited to citizens of the NATO nations.

The content of this publication has been reproduced
directly from material supplied by AGARD or the authors.

Published December 1981

ISBN 92-835-0305-8

Printed by Technical Editing and Reproduction Ltd
Harford House, 7–9 Charlotte St, London, W1P 1HD

THEME

The theme of this Specialists' Meeting was two-fold. It was first intended to make a retrospective judgement of significant developments which have occurred in the aerospace and defence information field, in particular, and the information science field in general, over the past two decades. Specific problems which to date remain unsolved (such as automatic indexing, fact retrieval, and input standardization) were reviewed. Then, the Meeting looked ahead to the '80's, speculating as to the technical and sociological changes which might occur, and trying to assess what will be the impact of these changes upon the information services available towards the end of the decade. The programme concluded with a user-supplier dialogue in the context of a general Forum Discussion.

THEME

Cette réunion de Spécialistes s'est organisée autour d'un double thème. Il fut tout d'abord procédé à un jugement rétrospectif des développements importants intervenus au cours des deux dernières décades, en particulier dans le domaine de l'information touchant la défense et les sciences aérospatiales, et, en général, dans le domaine de la science de l'information. Certains problèmes spécifiques n'ayant pas encore trouvé leur solution (comme l'indexation automatique, l'extraction des faits et la standardisation des entrées) furent passés en revue. Puis la réunion se tournait vers l'horizon 80, se livrait à des spéculations sur les modifications techniques et sociologiques susceptibles de se produire, et s'efforcait d'évaluer l'impact de ces modifications sur les services informatiques qui existeront vers la fin de la décade. Le programme s'achèvait par un dialogue entre fournisseur et usager dans le contexte d'un forum géneral.

TECHNICAL INFORMATION PANEL

CHAIRMAN: Mr H.E.Sauter
Administrator
Defense Technical Information Center
Defense Logistics Agency
Cameron Station
Alexandria, VA 22314
USA

DEPUTY CHAIRMAN: Mr H.K.Krog
Managing Director
Norwegian Centre for Informatics
Forskningsveien 1
Oslo 3
Norway

PROGRAMME COMMITTEE

Mr H.E.Sauter
(address above)

Mr R.Bernhardt
Leiter der Hauptabteilung Datenverarbeitung
Gesellschaft für Information und Dokumentation (GID)
Herriotstrasse 5
6000 Frankfurt am Main 71
Fed. Rep. of Germany

Mr J.G.Coyne
Manager
Technical Information Centre
Department of Energy
P.O.Box 62
Oak Ridge, Tennessee 37830
USA

Mr D.W.Goode
Chief Librarian
Royal Aircraft Establishment
Procurement Executive, MOD
Farnborough, Hants GU14 6TD
England

Mr H.K.Krog
(address above)

Ing. Gen. J.A.J. Guilleminet
Directeur Adjoint du CEDOCAR
26 Bld Victor
75996 Paris Armées
France

Mr A.Bruce
Head
Defence Research Information Centre
Station Square House
St Mary Cray
Orpington, Kent BR5 3RE
England

Dipl.-Ing. K.O.Hansen
Regierungsdirektor
Dokumentationszentrum Bw
Friedrich-Ebert Allee 34
5300 Bonn 1
Fed. Rep. of Germany

MEETING COORDINATOR

Dipl.-Ing. K.O.Hansen
Regierungsdirektor
Dokumentationszentrum Bw
Friedrich-Ebert Allee 34
5300 Bonn 1
Fed. Rep. of Germany

PANEL EXECUTIVE

Mr E.T.Sharp
AGARD/NATO
7 rue Ancelle
92200 Neuilly sur Seine
France

The Technical Information Panel wishes to express its thanks to the German National Delegates to AGARD for the invitation to hold its 34th Panel Meeting in Munich, and for the personnel and facilities made available for this meeting.

CONTENTS

CHANGING INFORMATION SCENE: FROM THE TIP SPECIALISTS' MEETING 1968 IN MUNICH TO THE EIGHTIES

Mr R K H BREE
An der Herrenwiese 6
5330 Königswinter 41
Fed Rep of Germany

INTRODUCTORY REMARKS

The Programme Committee of AGARD TIP has had the courage to invite me to present the first Paper of this year's meeting. This means inviting a colleague who retired from his professional duties almost ten years ago. I take it that the idea to do so has been influenced by the fact that, thirteen years ago and exactly in the same place, I had terminated the 1968 TIP meeting by summarizing it.

Looking up the Proceedings of the 1968 meeting I found that, by whatever circumstances, this summary had not been included and kept there for eternity's sake, yet I remember very well my statement of that time. I expressed some concern about the fact that the papers presented and discussed had much to do with the technical aspects and organisation of information services and very little with the situation of the user, even though the meeting was supposed to amount to a "User and Supplier Dialogue on Storage and Retrieval of Information". As usual, the suppliers' problems had dominated the meeting - much to my regret - and I said so. The end of this year's meeting might teach us how far the scene has changed in this respect.

First of all I would like to thank AGARD TIP for the invitation, which, of course, proved to be quite irresistible to me, not only for the sheer flattery it meant, but because it offered once more the opportunity to be with this group. My own former longstanding membership in AGARD TIP meant so much to me in retrospect that you will, I hope, permit me a few remarks on this relationship.

It has been almost exactly 25 years since I accompanied Erich Pietsch to Brussels to attend my first meeting of AGARD. Erich Pietsch must be remembered as an outstanding personality in our professional field. He was at that time not only the director of the world-famous Gmelin Institute for Inorganic Chemistry but as well later on the most resolute promoter of the ideas of Alvin Weinberg in this country. The TIP meeting in 1956 was directed by Gene Jackson of the U.S. We newcomers were at once warmly accepted and the unusually cooperative atmosphere of the Panel struck me at once. It didn't merely express a certain mood: the substantial practical assistance given to the newly emerging German centre for aeronautical information proved this at once. Especially grateful we had to be to our British colleagues, Messrs Holloway, Haylor and Vessey. These colleagues and others of the Panel Members became in due time real personal friends to me. I would mention just two names for all of them: Jan Schuller from the Netherlands and Jean Klopp from France. Unforgettable, too, will always be Col Woodrow Dunlop, in his time head of ASTIA, the information agency of the U.S. Armed Forces. His most courageous approach in tackling the problems of storage and retrieval at that time has been for the rest of us a substantial benefit, as we learned much from ASTIA's experience. I believe, however, that his authorities at that time were not half as grateful to him for his courage of sticking his neck so far out as we quite honestly were.

During my better than fifteen years of membership of AGARD TIP I had much opportunity to compare its performance to that of many other international organisations. In my experience TIP was always one of the best organisations because of its smooth and efficient cooperation. May this friendly spirit in the Panel go on for many years still to come!

Now to my topic of today: I shall try and look at the changing information scene, with emphasis on storage and retrieval, since 1968. At that time Alvin Weinberg's appeal to his government for paying much greater attention to the potential value of information for the modern society and its development had borne fruit. He had postulated much higher efforts for the careful use and evaluation of the worldwide recorded scientific and technical experience. He not only succeeded with his own government. He found serious attention, too, within the entire industrial world. Generally, one became aware that this recorded experience presented a treasure and an inexhaustible resource which had to be handled as carefully as any other vital resource. This meant that it had

to be made accessible as widely as possible. The governments in other countries had learnt their lesson, too, even though they had to foot the bills for the ensuing developments. As far as suppliers were concerned, we seemed to be in 1968 right at the beginning of a fast-spreading boom of information activities.

The situation was indeed challenging. Computers offered new and promising opportunities for storage and retrieval. Only a few information systems were at that time already fully operational and, if I remember correctly, all of them worked in batch-processing. Interconnections between operating systems were non-existent. In general, most of the operational steps for storage and retrieval were still in a testing stage and were much debated. The hardware in use was not only widely varied but in a period of steep development to ever-higher capacity and performances. This demanded a continuous adaptation of the programming for the suppliers.

The basis for any successful retrieval was the quality of the input. How to achieve dependable recall and sufficient precision was at that time much discussed. How to handle less familiar languages, especially Eastern ones, started to worry the suppliers as a serious challenge. However, the main efforts concentrated on mastering all the techniques for assuring satisfactory output.

The social, legal, and economic consequences arising with and from the operation of the new information systems had not yet come to play a major role during discussions at that time.

Only part of the user potential had already access to the emerging new systems. The vast majority still relied on their traditional means for getting access to recorded data on literature and facts. They browsed secondary publications and their indexes, checked tables of contents, and looked up handbooks and printed tables. In many fields they could rely for their guidance on progress reports of often remarkably high standard, which offered an extremely valuable help for singling out works of quality and were also a very useful tool for maintaining current awareness of their readers. Where are they left today? All this had a high human immediateness but - alas! - these do-it-yourself methods became more and more time-consuming and less and less satisfactory with the steadily growing quantities of new literature published.

Small wonder that the users started looking forward hopefully to the promised land of a new information technology. Its prophets did much for stimulating high hopes, possibly even too high. I often had the feeling at that time that in the heat of the discussions the term "information" was too often (and in gross simplification) used for the output of the new systems, even though this output could not offer more than bibliographic data on available literature. This amounted to suggesting to the user that the new systems would be capable of relieving him fully of any further use of his brain. The new systems seemed to be able to spit out on demand and in no time all the relevant answers to the users' problems, ready for instant use. How misleading this concept was the users would soon come to experience.

For me, and in our context, the term "information" should cover that process which goes on inside the brain of the user when he is confronting his professional problem with the experience of other people, as recorded in whatsoever publication. At their very best good information systems can only support that process by providing data on relevant recorded experience, be it on facts, be it on literature. This contribution is today certainly indispensable in view of the quantity of reporting, but even so it can never be more than a preliminary step for the subsequent intellectual process. You know the old saying: "Beauty is in the eye of the beholder"? In a very similar way one could state: "Information is in the brain of the reader"!

Those using at that time the emerging new services soon came to recognize that the quality of the answers received from the systems depended mainly upon the quality of their own requests, - that is, the honest ones admitted this. For others, any less satisfactory results derived solely from the deficiencies of the new systems. Not all of the users were ready for the inevitable changing of their habits, when compiling their requirements.

In 1968 the ancient tradition of the libraries of giving free service from their holdings prevailed as well in the new information services. At its worst the users had to pay token charges.

I remember 1968 to be a time full of exploratory initiatives and high expectations on the side of the suppliers and full of suspense and hopes on the users' side.

Details of the present situation on the suppliers' side are much better known by this audience than by myself. The excellent contributions to be expected from Mr Ember, Monsieur Yanez and Mrs Lamvik are coming closest to depicting the present situation in its full complexity and from many more angles than it is possible for me to consider in this introductory contribution. I think we can agree that the most significant change since 1968 is the fact that, in the meantime, millions and millions of data on literature and facts on just about every field of human interest and activity have been stored and will be added to by millions of data year by year. The methods of retrieval are no longer a crucial problem. Batch processing has been replaced by on-line processing. On-line dialogue via terminal has become a standard procedure and modern communications-technology is providing means of direct access to almost all important data-bases and of interconnecting data-bases and users within large networks.

Not only has the quantity of stored data grown. So has the number of offered services and institutions. And the so-called "non-technical factors" are influencing the operations of the information systems extraordinarily. All these influences are carefully analyzed in the Paper of Monsieur Yanez, starting with the more operational ones and stepping over to the political and economic factors and the manifold legal considerations, without leaving out the thorny linguistic problems. Further constraints are added by the various communication-processes. In their cumulation these influences can build up high obstacles for the end-user. Any user looking for pertinent results on his requests, especially on multi-subject, problem-oriented ones, might feel himself placed right in the middle of an almost impenetrable thicket of offered services with differing rules and regulations, and with unknown overlaps and gaps between them. He can find the best way through this thicket only if and when he has the time and the will to become familiar with the complexity of the conditions which he is facing.

This is but one aspect of the changes on the users' side since 1968. Doubtless the number of users of information systems has multiplied. Those using these facilities regularly have adapted themselves much better to the conditions of their use. It would be very interesting to know at least roughly what proportion of the user-potential is using the information systems regularly and the existing differences from country to country.

However, I believe that in the most important respect the user has not changed at all. He has remained basically what he was in 1968: a single human being, eager to solve his professional problems and eager to support this effort by getting himself informed as well as possible, but - alas! - with the same limited personal capacity and limited time for digesting and evaluating informative data and texts. In view of the tremendous development on the suppliers' side in the quantity and diversity of services and holdings he is much worse off than ever. Even those conscious of all the advantages of being fully informed: can they afford to devote an ever higher proportion of their professional time to singling out and digesting what would be helpful for them? I am afraid the structure of the offered services is too complex. Rarely is one single supplier in a position to answer exhaustively any problem-oriented question. Yet, even incomplete services are asking for money and the user has to convince his boss that this money will be well spent.

This rather glum situation for the user demands a way out and there seems to exist a helpful solution: enlisting the services of an intermediary, an information-broker. By so doing the user himself can concentrate on formulating his requirements. His partner will - in the ideal case - unburden him from the choice of the most suitable information services, from scanning the results of the retrieval, extracting the needed data, from supplying the texts the user might wish to read and, if need should be, from supplying the necessary translations.

In larger industrial firms, research institutions, and administrative units users can almost always rely upon this kind of assistance from their colleagues in the in-house libraries or information departments. These helpers not only understand the problems and needs of their user-group but have also become very knowledgeable on the different information services and their specific properties. It seems to be this group of users which has installed terminals and direct access to existing networks. But only some of the users are in this privileged position. Those in smaller and medium organisations, not to speak of individual users, are in a less favourable position. There seems to be a trend, which is to close this gap: professional societies, special libraries, and as well chambers of commerce, are offering their own particular services and the demand seems to be already sufficiently well developed for free-lance information-brokers to enter this field too.

I might be mistaken but I think that this development is very promising. It relieves the end-user from the task of having to become familiar with the often not-so-transparent structure of the offered services. He can concentrate on his real task without having to forego all the stimulation the confrontation with recorded experiences can mean to him.

There are a few more aspects within the changing scene since 1968 I would like to turn to. Alvin Weinberg, in his historical report, set much store on the quality aspect and, quite logically, exhorted authors to maintain as high a standard as possible in their presentations. I think all of us are familiar with the strange situation, that those users requiring information are much more demanding than they used to be, when they are authors themselves. In retrospect, it can certainly be stated that Weinberg's postulates were absolutely justified but that the majority of authors did not act accordingly. The consequence is that this slightly malicious remark of Gene Garfield for characterising the new information systems seems still to be valid: "garbage in, garbage out!". I am afraid there is no easy way out. Preselection before input means an invitation to unlimited discrimination. Consequently, the data bases are full of data on less valuable literature. Unfortunately, the costs of storing these data don't differ from those of storing data on valuable records. The next consequence is that, at the user's request, data will be furnished on relevant but worthless publications. Weinberg had recommended the setting up of special analysis centres. They were to have been manned by highly knowledgeable specialists able to exercise their judgement so as to single out the best portions of the publications. I don't know whether or not any such institution has been created as a public service.

With regard to the quality aspect it is perhaps worth mentioning that in very many cases the user is no longer able to judge the quality and validity of the service of the supplier. The only thing he can be sure of is that he has to pay for the services received, - quite independent of the true value of the service received. This is a strange situation which could be tolerable only insofar as the supplier is fully conscious of his responsibility, that is, when he accepts the position of trustee via-a-vis the user. However, is this position as a trustee compatible with the financial constraints of any supplier who has either to make money or at least to try to break even? It would be so easy to save on the back of the paying user, when one can hardly get caught red-handed. I am far from saying that this actually happens but is it to be fully excluded?

In this same context, every once in a while nowadays the term "information-market" is mentioned, maybe intentionally, for describing the supplier-user relationship. This would, of course, suggest that there exists a market with rules of supply and demand and with a healthy competition between the suppliers as a regulatory element. I am afraid that such a market does not exist in reality. On the suppliers' side we are facing, for the most important subject fields, de-facto monopolies. Nobody can afford to double up INIS, nor INSPEC, nor CAS, nor any other of the outstanding services, and that only for the sake of competition. The financial and intellectual effort would be too big and as well too risky. So we must wonder: where are then the regulatory forces of the market which are supposed to protect the users' interest?

Parallel to this subject monopoly, which is as significant to me as it is inevitable, we are facing as well a kind of linguistic monopoly, again an inevitable one, but with consequences for important user-groups. It is a fact, just a part of the history of our profession, that most of the operational databases originated in English-speaking countries. This is quite understandable because here was located the highest user potential. Even though we have to accept this fact it is nevertheless creating problems for the supply to the non-English world and its users. Furthermore, I am afraid that this situation might lead - again quite understandably - to a relative under-representation of the literature in languages other than English. Doubtless this would mean a discriminating effect, even though it might happen quite unintentionally, if not dictated by consideration of the much higher cost for the input operation. I feel it is regrettable that there are not more databases organized like INIS or AGRIS, where equal participation is the basis of all operations. But it is certainly much too late to expect any further developments of this kind.

From these considerations it is only one step to the fact that national independence in access to recorded scientific and technical data simply does not exist for a large number of industrialized countries. The artificial creation of national networks does not change this situation. Moreover, these national networks do already, or will in the near future, compete with international commercial networks. Coming back once more to the idea of Alvin Weinberg: he likened information to other essential and ind. ensable ources - and rightly so. But this might mean that during the Eighties we w. ome face a similar situation in our field as already exists with other resour , like energy, raw materials, etc.

What this could mean in the long run remains to be experienced. At worst it could mean stepping away from the centuries-old tradition of free exchange of ideas, theories, and experiences in science and knowledge. Together with the trend to commercialisation of information services it could very well mean that even something as sensitive as information handling might become more and more subject to all the constraints of commercial survival. This would put all the legitimate expectations of the users at the mercy of forces outside of their interests.

Unfortunately, users do not constitute a pressure group. Unlike customers in any normal trade they cannot even use the weapons of refusal. Is there any counterbalance to these rather adverse possibilities? Personally, I believe that, so long as we are willing to heed the ideas of Alvin Weinberg and to maintain the outstanding place he gave to the importance of this world-wide treasure of recorded experience, as long as we continue to think as he did, we cannot refuse nor play down the corresponding serious responsibilities on the suppliers' side. It might very well be that in the Eighties many suppliers will become torn between this professional responsibility and growing financial or commercial pressures. This problem might become a steady companion to any supplier.

As regards the users, I do hope that in the Eighties the manifold attempts at giving higher transparency to the nature and content of the offered services might succeed - together with the hopefully started efforts towards a broader standardisation of the rules and regulations for the use of the offered services. Furthermore, I am hoping that fair international cooperation can go on and does not become victim of commercial competition. I don't think that I have hidden my own old-fashioned concept, that information handling is and should remain a task as indispensable as it should be conducted responsibly. Therefore I consider it a danger to act in this field in the same way as if one were trading in any whatsoever commodity. I think all of us need to do our best to try to develop the art of serving our users in the best possible manner, if only in the interest of contributing to the survival of our society.

Major Developments in Modern Information Services: the Canadian Experience

George Ember
Canada Institute for Scientific and Technical Information
National Research Council of Canada
Ottawa, Canada, K1A 0S2

Since the last TIP Specialists' meeting in 1968 technological progress dramatically changed the setting and the capabilities of information services. The routine patterns of data storage and retrieval manipulations have been fundamentally restructured and the interface between system and user is now bridged over by powerful linking mechanisms. Most of the major developments of information services in the intervening period can be traced to

1. advances in computer design and automation,
2. improvements in telecommunication and
3. the union of computer and telecommunication technologies.

Computers increased their storage capacities, core memory potential, processing speed, while their size, cost and operating expenses dropped significantly. Telecommunication introduced new modes of digital data transfer, whose cost is less dependent on geographic distance, and is now developing inter-city and inter-continental connections via inexpensive fibre material. The joint power of computers and telecommunication maintains various modes of on-line services, links remote processing hardwares, supports networks and provides global dimensions for the information industry.

The consequences of these developments dominate the modern service environment where the user assumed a command position over electronic machinery and telecommunication facilities. Perhaps the most notable character of modern information services is the user's participatory role and his changed attitude toward the information-gathering activity which was once time-consuming and cumbersome and became easy and convenient. Either directly, or through an intermediary, the user can now plug into an international electronic marketplace where information, as a commodity, is in abundant supply and readily available for his individual use.

It is also of importance that the increasing complexity of services are usually not accompanied by an increase of sophistication in using them. The rule almost generally applies that the more complex a system is, the easier its user-interface can be managed. Signing-on procedures, command languages, file interrogation techniques and various other manipulations can normally be learnt in a few training sessions.

The large variety in the service capabilities of modern information systems compels me to select carefully only a few characteristics and features which qualify as "major developments". I have chosen the following four properties as the primary potentials of modern information services:

1. Searchability of bibliographic records of the world's scientific and technical literature.
2. Access to critically evaluated numeric data and research results.
3. Ability to bring together people to solve problems and jointly develop ideas.
4. Means to locate source literature in full length and reliable document delivery.

While reviewing briefly these properties, I shall refer to the Canadian experience in scientific and technical information services only to illustrate the discussed operational routines.

1. Searchability of bibliographic records of the world's scientific and technical literature

According to the 1981 Aslib directory of "Online Bibliographic Databases" by Hall and Brown, there are now some 70 million bibliographic references searchable via computer and over 85% of these are in the natural sciences and technology.

Data files covering entire disciplinary fields (chemistry, health sciences, etc.) or mission-oriented research areas (oil sands technology, occupational safety), specific types of publications (technical reports, patents), represent the main body of data which have been made searchable in the on-line and off-line modes. Data bases which are commercially accessible through the vendors of the information industry display the literature of the last 10-15 years. This is sufficient to cover the time span of the half-life of current scientific and technical knowledge.

Beyond the publicly available stock of retrospective files, there is a very large number of private data bases of research workers and institutions. With the exception of the

proprietary files of industrial research establishments, some of these private data bases are often, or become eventually, available to a professional community or a national clientele.

The most popular international data banks are electronically stored cumulated versions of printed indexes and abstract journals. They can be scanned by on-line searching or used to update individuals recurrently through the selective dissemination of personalized sets of citations.

In Canada bibliographic search services are offered by the private and public sectors of our own information industry and by international vendors. One of our on-line bibliographic search system is CAN/OLE (Canadian On-Line Enquiry) which is operated by the Canada Institute for Scientific and Technical Information (CISTI), a division of the National Research Council of Canada. CAN/OLE, which has been in existence since 1974, now offers 18 files for searching via some 700 terminals across the country. The files contain about 12 million citations and the system was utilized for over 6000 connect hours in the past year. The Canadian MEDLINE network operates through 110 centres at the present time and supports 17 files. These files are not mounted in Canada but can be accessed via direct hook-up with the MEDLARS facility of the U.S. National Library of Medicine in Washington. The updating portions of the CAN/OLE files and the monthly supplements to the MEDLINE data bases are also searched in batch mode by CISTI's Selective Dissemination of Information service (CAN/SDI) which has been in operation since 1969. CAN/SDI now serves roughly 2200 subscribers and scans over 3000 individual interest profiles for its clientele.

In Canada, as in any industrialized country, bibliographic search services are instrumental in the promotion of economic and social development. Canada produces only about 3-4% of the world's scientific and technical literature and is a "net importer" of information. With a very few exceptions all countries belong to the same category and their science and technology must tap the pooled worldwide resources of published knowledge. It is certainly one of the major developments of modern information services that now they are able to accomplish this essential task.

2. Access to critically evaluated numeric data and research results

Scientific research is a prolific generator and constant user of measurements. The basic and applied scientist or the engineer must often consult elaborate tables or graphic representations of quantities and values, handbooks, laboratory manuals and the serial literature to find facts in the form of analyzed and evaluated numeric data. A relatively recent development is the compilation of numeric data in data bases which can serve investigators as research tools capable of two essential functions.

Mounted on computers, numeric data bases not only offer "look-up information" by locating and retrieving items of interest from the content of regularly updated files. The selected set of data can also be manipulated at will in a wide range of theoretical and experimental applications. They have a high degree of maneuverability and graphic capabilities further support their versatile use.

It is expensive and very laborious to create and maintain numeric files and rigorously assure their utmost accuracy. Despite their relative scarcity at the present time, files are being developed in growing numbers in the United States, United Kingdom, Federal Republic of Germany, France, Japan and elsewhere.

To exemplify the complexity of such a file, let me point to one data base that originates in the Crystallographic Data Center, Cambridge, England, and contains the results of some 28,000 independent studies of 25,600 chemical compounds. This file grows at the average rate of 3000 new entries annually. The three linked sub-files of the Cambridge Crystallographic Data Base contain bibliographic data (name, formula and class of the compound and literature reference), connectivity data (chemical connectivity of atoms, bond types, etc.) and, finally, comprehensive data on the properties of the compound (unit cell, symmetry, atomic co-ordinates, bond length, etc.).

So far 5 Canadian numeric data bases have been developed in the fields of chemical engineering, chemical thermodynamics, thermochemistry, electrical power engineering and solid engineering.

Since 1978 CISTI has made accessible in Canada the SPIR (Search Program for Infrared Spectra) data base with some 143,000 entries. The system enables the user to locate the spectrum closest to his unknown. CRYSTOR, the Cambridge Crystallographic Data Base, which I mentioned earlier, became a national service last summer and is now in full operation in the on-line mode.

The availability of manipulative numeric data files, with their specially designed software, should be considered another major development whose impact is already visible. It is expected that in the coming years this variant of modern information services will expand its sphere of applicability and become one of the common tools of researchers in scientific and technical disciplines.

3. Ability to bring together people to solve problems and jointly develop ideas

National inventories of ongoing research activities, directories of projects subsidized by governmental granting agencies and private foundations, selective listings of experts in various scientific and technical specialties, and other data cumulations of this kind, have many uses in policy planning, scientific management, in monitoring the directional trends of scientific communities, etc. They also fulfill an important function for the working scientist and technologist. The potential benefits of this function (which is usually defined as "referral") are now implemented with demonstrated success.

Scientists often need intellectual help that only other scientists can offer. In some cases this help is not available in the immediate environment of the investigator and the literature does not contain the clues which can lead to the solution of his problem. Science and technology has strong communal characteristics, a nationally and internationally scattered population that is stratified and contained in "invisible colleges" and various types of insular settings.

There is a class of information which can be found only in the laboratory, close to the workbench of the researcher. Modern information services can tell a great deal to their users of "who is doing what and where", "what is the status of the activities of others", and "who are the experts of an investigative topic". The basic instruments for providing such information are:

1. research inventories (e.g., SSIE in the USA, IEC in Canada),

2. data banks for the "fugitive literature" (e.g., NTIS for technical reports, data bases for conference presentations) and

3. data banks on expertise (e.g., KSI in Canada).

Research inventories in machine-readable form, usually with hard copy or microform versions, are descriptive records of individual projects which identify principal investigators and co-workers by name and affiliation. To some record formats a string of keywords or descriptors from a controlled vocabulary, and often abstracts, are added and also locating information concerning the availability of progress and final reports generated by the project.

Variations in scope and comprehensiveness range from narrow sectorial compilations (such as Arctic research) to an all-inclusive, retrospective and minutely annotated search system which is maintained by the Smithsonian Science Information Exchange. The Canadian IEC (Information Exchange Centre) file is limited to federally supported academic research projects since 1970 with an annual growth of about 10,000 entries.

The "fugitive literature" of technical reports display scientific and technical activities in the close-up dimensions of magnified details. Their elusiveness is due to the limited circulation of accounts on governmental in-house projects, contracted research and activities with grant support. Access to U.S. report material via on-line search, and in microfiche and hard copy form, through the National Technical Information Service (NTIS) is invaluable for the working scientist and engineer. Other searchable sources of smaller proportions, such as the CODOC data base of the CAN/OLE system in Canada, are further examples of this service. Papers presented at conferences and translation of research articles from foreign languages, which also belong to the fugitive literature, have become more manageable in recent years by ISI in the U.S. and a number of other data services.

Data banks containing records of subject experts, with well-defined indicators of their specialty, are often not made public because this service must be carefully protected from abuse. The task of this sensitive utility is to promote direct personal contacts, consultative links and partnership of people and all the related functions are dependent on the consent and circumstances of individuals. The Knowledge Source Index (KSI) with the detailed and continuously updated records of some 3500 Canadian experts in government, universities and industry, is one of the examples of this information service.

Creating awareness of the working interests, accomplishments and the know-how of members of professional communities has far-reaching significance in the advancement of scientific and technical knowedge and practice. Effectively used, this type of service has great potentials in the team-formation and personal orientation of individuals and, therefore, deserves the rank of "major information services".

4. Means to locate source literature in full length and reliable document delivery

The 70 million bibliographic citations displayed by on-line services recommend reading material to the user to satisfy his information need. The result of a successful bibliographic search is, therefore, the first action in a chain of events which leads to the acquisition and use of the information. The three further steps, following the bibliographic identification, are:

1. location of the cited item,

2. requesting the material from a source. and

3. delivery of the document.

Machine-readable catalogues of monographic and serial collections, automated union lists, became standard sources for location information in the past decade. Computer-produced microfiche (COM) versions of these tools and conventional hard-copy volumes are also widely used.

Information systems often support the location search by including source data in the citation format or by mounting catalogues and union lists on the same computer system which provides the bibliographic files for search. After completing the search manipulations, the user can switch to a national holding file on the same system to find immediately a source for the photocopy he wants to order. During this maneuvre he can find out whether his own library holds the required item or, if not, which is the holding library closest to him.

In Canada the UTLAS (University of Toronto Library Automation System) maintains a catalogue of the major academic and governmental collections in searchable machine-readable form. A Canadian version of the DOBIS library management system of Dortmund University in the Federal Republic of Germany was jointly developed by the National Library and CISTI for shared cataloguing functions among its fields of application. CISTI's on-line information system, CAN/OLE, includes the national union list or scientific and technical serials in a set of 18 data bases, along with detailed files on the document collections of two important governmental libraries: the departments of Transport and Environment.

On-line document ordering and the so-called electronic mailbox facility were designed to automate the document requesting phase. This development took place in the past 5-6 years and since then became a component of large North American and European information networks.

CISTI backs up the CAN/OLE, MEDLINE and CAN/SDI services with photocopies from its serial collection of 32,000 titles. Via on-line ordering, telex and mail, 184,000 requests were sent to us last year. Of this volume of requests 76% was filled from CISTI's own collection and another 20% were forwarded to the other 253 scientific and technical libraries whose holdings are included in the union list.

Document delivery, the final step in the chain of events, is still carried out by conventionally transporting the item from the source library to the user. While holding information can be obtained by automatic means and the ordering routine for photocopies and loans has been enriched with automated elements, we cannot report at this time major developments in document delivery. Telefacsimile transmission of full text has not been employed on the large scale and videotext technology is still a promise for the future.

The only Canadian innovation in this area is the bulk delivery of items from CISTI in Ottawa to distant cities. The combination of courrier service and air transport carries document packages to drop-off points from where they are locally distributed to their final destination. Considering that about 80% of all the requests filled by CISTI have a 24-hour internal turn-around time, and bulk delivery speeded up considerably the transportation procedures, the time frame of this transaction was reduced to the minimum under the present circumstances.

*

I have attempted to review the major developments in modern information services with examples taken from the Canadian experience. The review I have presented was very selective and focussed on the user-related operational nature of services, leaving essential related topics, technical issues, managerial and economic problem areas, projections for the future, to other presentations in the program.

We live in an age of constant developments and our professional environment, our tools and practices, are always transitional. Whenever we meet there are new major developments to talk about - and this is the most stimulating aspect of our profession.

MAJOR DEVELOPMENTS IN HARDWARE, SOFTWARE AND TELECOMMUNICATIONS IN THE FIELD OF INFORMATION STORAGE AND RETRIEVAL

by

Dr-Ing. D.Morgenstern
Siemens AG
DV System Technik DB/DC
Hofmannstrasse 51
8000 Munich 70, Fed Rep of Germany

Summary:

The driving force behind the growth of data processing up till now has been the explosive development of technology. The availability of semiconductor chips with an ever-increasing component density has made a continuous improvement in the price-performance ratio of computer systems possible.

During the seventies the mainframe power of an average dp-installation increased at least tenfold, the main memory size by a factor of 30 and disk memory by a factor of 100. Over the same period floor space, power consumption and costs decreased.

The utilization of dp-systems was improved by innovations in systems architecture and software. Typical examples of this are virtual addressing, microprogramming, improved protection mechanisms and data base systems.

The use of telecommunications in computing applications has expanded rapidly. Between 1970 and 1980 the number of terminals in use in the Federal Republic of Germany increased by a factor of 45, similar increases are also common in other countries. During the same period data networks for public use appeared in all major western countries.

Due to differences in telecommunications law, data networks in the United States and Europe evolved in different ways. Standardization activities in this field will form a basis for generally available public data networks in the near future. During the eighties low cost connections to data networks will be made available to the private citizen.
At the end of the sixties information storage and retrieval systems were only available in a rudimentary form, today however quite sophisticated real-time information systems are employed. As a result of the progress made in dialog-processing and access methods for non-specialized users information systems have enjoyed increasing popularity.

At present research is being carried out on natural language text processing.

In the future character recognition, speech processing, intelligent terminal systems and public data networks will be the center of interest.

1. Hardware developments

Three 3 key factors have contributed to the recent development of large CPUs :-

Progress in semiconductor technology between 1970 and 1980 has made it possible to increase the number of logic functions implemented on a silicon chip by a factor of 100.

Semiconductor memories with an access time of less than 10 nsec are available.

Techniques which permit the packing of a large number of silicon chips in a very small volume have been developed

This means that CPUs capable of carrying out several million operations per second can be constructed whereas the present costs are only a fraction of the costs involved ten years ago.

Since 1970 main memories have generally been implemented using semiconductor memory and capacities of several million characters are quite common. Disk memory, with capacities of a few hundred million characters per disk drive, has been developed and used as secondary memory. Attempts to find a memory accessible by purely electronic methods, e.g. bubble memory, and so replace disk memories which operate on an electromechanical basis, have not been successful. Development work on magnetic tape memory has been very tardy and apart from improving the recording density by a factor of 4 no further innovations have been made.

In the seventies there were several attempts to market electro-mechanical tape library systems, but none of these were resoundingly successful.

A landmark, as far as output devices are concerned, is the appearance of the first electro-photographic high performance printer on the market. These devices operate using a laser printer system. Apart from a very high printing speed, they offer character sets which can be exchanged electronically and graphics. The matrix printer is by far the most popular type of slow printer in the lower price range. The ink jet printer is another type of printer which has been developed in the meantime. It operates silently and many are capable of producing printouts in several colours.

The microprocessor, whose capacity and complexity is continually increasing has been and will continue to be a key factor in the development of hardware for terminals and small computers.

2. Architecture and software

Looking back over the last 10 years it is obvious that innovations in computer architecture are just as important as the technological developments that have taken place.

At the beginning of the seventies it could already be foreseen that the increasing complexity of data processing systems would, sooner or later, cause serious reliability problems. The large number of components constituting a data processing installation and the certainty that this number would continue to increase was a reason for fearing that random faults would occur so frequently that the time between breakdowns would only be a few hours. This problem has however been solved by means of automatic error correcting procedures. Today, for example, errors in the main memory are usually corrected automatically.

In the past computer commands were often implemented in the hardware of the device, nowadays microprogramming means that an increasing number of more powerful commands can be added to the computer's repertoire.

Two main advantages resulted from the increase in main memory size - increasingly larger and more complex programs could be run and computer time could be shared simultaneously by a greater number of users. This meant that the danger of an error in one program causing other programs to malfunction became more acute. Virtual addressing and protection mechanisms were introduced to eliminate these problems. In this way the access of individual users to data and programs could be checked.

Multiprocessor systems were created when the demands for more computer power could no longer be satisfied by the largest processors. This involved distributing the system load among several processors which operate together in a well defined way. The main memory of the system was also replaced by a memory hierarchy consisting of a small fast memory and a very large slow one. These changes resulted in a marked increase in operating speed.

In the early days of data processing the manufacturer delivered a very simple operating system, which realized a few basic functions, in addition to the hardware. Since then system programs with several million commands have been developed. Today many of the functions the user requires are already provided by the operating system which has standard interfaces. This means that the user no longer has to spend as much time as he did before writing his own programs.

At first it was feared that the increasing number of elements in operating system codes would cause an increasing program error rate and consequently lower stability and reliability. In the meantime however new procedures for designing and implementing software systems were developed in a new branch of applied science called software engineering. The result of these new procedures was increased software quality and performance. Great advances have been made in data security procedures and methods of restarting the system when faults occur.

The costs involved in system software development have, in the course of time, become considerably larger than those of hardware development. For this reason payment in advance for system software is a condition imposed by all system suppliers. This led to the creation of a software market with strong competition between numerous software houses and the system suppliers. Today the user can buy ready made software packages to solve many of his software problems.

3. Telecommunications and Data Networks

Teleprocessing has existed since the inception of data processing approximately 30 years ago. The main function of teleprocessing then was allowing the user to carry out input/output operations at a location remote from the computer centre. Remote batch stations connected to dedicated lines were used for this purpose. In the late sixties video units became available and made dialog between user and system possible, this dialog was frequently effected via the telephone network. At first, because of technical limitations, the lines were connected to the CPU by means of hard-wired control units. Teleprocessing accessing methods were incorporated in the operating system to make a few basic functions available to the user. The first time-sharing operating systems were introduced around 1965. Time-sharing is the allocation of a certain amount of CP time to each user in turn. This means that, in effect, each user has the complete system at his disposal. This mode of operation is especially suitable for interactive programming and scientific applications where each user is executing his own specialized jobs. This is, however, not the case for commercial applications where many users want to run the same programs and access the same data more or less simultaneously. Computerized booking systems or systems for accessing shared data bases are applications of this kind. Clearly time-sharing is not appropriate and the approach adopted must be the serialization of individual jobs by special system software components. In 1970 the first TP monitors wete developed to solve the serialization problem as well as other problems relating to safeguarding and program restart. This made transaction processing possible and after 1975 the most important

transaction functions became available in operating systems. Since then data processing has been applied to more and more areas of commerce where a large quantity of similar, repetitive tasks have to be carried out. Over the same period of time the number of dedicated lines increased so rapidly that each group of lines had to be controlled by a satellite computer.

Private data networks were set up shortly afterwards using dedicated lines and concentrators, today for example, large airlines have networks interconnecting thousands of terminals all over the world. The telecommunications laws in Europe and the United States are very different and this is why the data networks in both countries have followed different lines of development. In Europe the administrations have a telecommunications monopoly as private citizens are prevented by law from setting up telecommunication systems. This means that the development of data processing in Europe was heavily dependent on the development of public networks. West Germany is typical in this respect as over 50% of terminals are operated via the public telephone network. Approximately 40% use a public dedicated line network which was set up by the telecommunications administration in the early seventies when the rapid growth of private line communication systems threatened the state monopoly. A new digital dialing network (DATEX L) has also been set up recently. A milestone in the development of public networks was marked by the introduction of EURONET at the end of the last decade. In 1980 a public packet switched network (Datex P) went into trial operation in West Germany. Table 1 gives an overview of the increasing use of data terminals in the various public services in West Germany.

Table 1

The number of terminals using various West German communication networks (Source: Deutsche Bundespost)

	1967	1970	1975	1980
Telephone networks	170	1407	13989	39149
Dedicated line network	-	-	14168	60562
Datex network	60	404	1419	3548
other networks	97	2447	7770	4234
Total number:	327	4258	37346	107493

In the United States teleprocessing networks operated by the private sector and available for general use were introduced a long time ago. APRANET, TELENET and TYMNET are typical systems of this kind. New techniques such as packet switching were tried out in these networks which, as a rule, were only available to certain user groups - research institutes or the customers of service computer centres for example. Generally speaking these networks did not offer a complete service as far as area coverage was concerned. However, because the market for high performance data processing has grown so rapidly in recent years, AR & T and other large organizations are considering setting up data processing networks with full area coverage. As well as the transfer and storage of data these networks will also carry out a certain amount of data processing. Attempts are also being made to set up teleprocessing networks using satellites. The main characteristic of such systems is their high data transmission rate, which also makes other services such as picture transmission or video conferences feasible. Because the cost of satellite ground stations is still relatively high, satellite networks will only be accessible to users with considerable resources at their disposal.

During the seventies the developers of data networks in Europe lost ground to their American counterparts because they were prevented from setting up private networks. However, as the eighties approached, the situation changed as the techniques used in digital dialing networks had reached such a high degree of development that they could be made subject to international standards. After the CCITT published the X-series recommendations work started on the construction of public digital dialing networks in most European countries. In the years to come these networks will offer a range of remarkable new services, which could also be important for information and documentation centres:

a) TELETEX offers a new kind of telex service with a high transmission rate and a large character set. This service is primarily intended for text communciation and office automation, it will probably, to some extent, replace the existing telex service. The result of this will be the development of a new type of intelligent printer terminal, which can also be used in teleprocessing

b) PRESTEL, the TELETEX system originally developed in Britain, is now being tested in several European countries. In this system inexpensive user terminals are connected via the existing telephone network. Teletex will be standardized in the near future and is particularly well suited for dialog with data bases and for booking and ordering services. It can also provide full area coverage with transaction processing. Experts predict that travel bookings, mail order purchases as well as other financial transactions will all be possible from one's own home in a few years time. The replacement of telephone directories by very low cost terminals employing a retrieval system is an idea originating from the French Telecommunications Authorities which is worth mentioning.

c) The first fully digital telephone exchanges have been operating in Great Britain since 1980 (System X). In addition to digital telephone traffic the exchanges can also carry out teleprocessing. Agreement about the standardization of European digital telephone systems has almost been reached in discussions at CEPT. Using this system it is envisaged that speech communication will take place at 64 Kbit/s., digital traffic is also possible as is the transition to other public networks. The first prototypes using these new techniques are expected to go into operation around 1983. It is probable that several European telecommunication authorities will start introducing Integrated Services Digital Networks (ISDN) in 1985

d) Experiments on wide band communication using co-axial cables or optical fibres have been carried out all over Europe. These systems have a large bandwidth but are in general unsatisfactory for dialog traffic, as they usually have a tree structure.

e) In the latter half of this decade satellite communications will become increasingly important in Europe. Telecommunications authorities are also considering the possibility of making wideband services available to the private user.

In conclusion one can say that after several years of divergent development in Europe and the United States, the European authorities have become very active in the field of public data networks. This would lead one to expect that users will soon be able to benefit from the advantages of public networks, i.e. full area coverage, guaranteed quality of service and standard interfaces.

The variety of digital networks and transmission procedures has considerably increased the complexity of hardware and software. Also, because of the rapid development of private networks in the USA in the mid-seventies, almost all the important manufacturers of data processing equipment have their own "network architectures", which are, in general, not mutually compatible. The ISO has been trying to gain acceptance for the standardization of a reference model, which describes the functions of teleprocessing systems in terms of function layers. CEPT, CCITT and ISO have begun to publish recommendations and standards for teleprocessing systems based on this model and it is hoped that this will improve the compatibility of data processing systems in the years to come. No technological advance mentioned previously has greater significance for teleprocessing than the microprocessor. Without the microprocessor it would not have been possible to construct intelligent terminals which have all the characteristics of a small computer system. These devices are very common today in personal computers and the number of applications is increasing rapidly. In the future they will be equipped with their own small data bases opening up further possibilities for the user.

4. Information Storage and Retrieval

At the end of the sixties information storage and retrieval systems were still in a rudimentary stage of development. Higher functions could not be realized as the required CPU speeds were not available and economic problems often arose because memory costs were high. However the situation has changed dramatically in the course of the last 10 years. Modern data processing systems can effectively store a very large amount of data and are able to accommodate a large number of users simultaneously.

In the seventies the development of information storage and retrieval was determined by three factors:

1) - Batch processing in a computer centre was the usual mode of operation in the sixties. However 1970 saw a breakthrough in the form of on-line systems which could be used by many users in dialog operation.

2) - The advances made in teleprocessing and network design led to the formation of extensive information networks such as ARPANET, TYMNET or EURONET. In their turn these developments stimulated the construction of new, public switched data networks in Europe.

3) - While large data bases were being developed and set up for use in many areas of science and technology, an analogous application of teleprocessing to information storage and retrieval was also taking place.

To recapitulate we can say that information storage and retrieval systems have become commercially viable in recent years and that an information industry has been created which will be probably be a sine qua non for future economic development.

Most commercial information storage and retrieval systems in operation today are designed using principles that were already familiar 10 years ago. This means that practically all storage and retrieval algorithms are based on Boolean logic. The conclusion that no radical change has taken place in this field over this period is inescapable. Mathematical procedures for measuring the similarity of data had already been proposed at the beginning of the seventies, however they could not be realized in practice because the demands they made on computer resources were too great. The increasing quantity of documentation which has to be stored in modern retrieval systems is one cogent reason for assuming that these methods will not find much favour in the near future.

The automatic indexing of information has not proved to be commercially viable even though several effective programs have been developed. On the other hand a commercial breakthrough in the near future is likely.

Linguistic data processing methods have been implemented in experimental systems, but as yet no wide practical application for them has been found. Nevertheless proofs exist showing what the exact limits of such methods are. Methods of evaluating the semantic content of information will not be available in the foreseeable future.

The use of commercial data base systems has grown rapidly in the seventies and to-day the use of data bases is common practice. Over the last 10 years progress has resulted from research into various data models, of which the relational model has made the greatest impression. The main area of debate for experts in the field has been the standardization of data base interfaces. CODASYL was the starting point for the discussion and is supposed to facilitate the portability of user programmes for various data base systems. It is worth noting that standardization was limited to the data base systems as the user interfaces of information storage and retrieval systems are, in general, used in processing dialog operation and this meant that the standardization of programming interfaces did not seem to have the same importance. New developments in data base systems were also accompanied by greater ease of use, providing numerous aids that make life simpler for the system user whether he is an application programmer or a dialog subscriber. These aids include interactive query languages, report generators and data dictionaries. Automated data base design has also simplified the complex problems encountered in this field.

Two main trends can be seen in the future development of data base technology - the implementation of extremely large data base systems and data base decentralization.

Even today the problems associated with "distributed data bases" have not been fully solved. There is a need for systems in which data is stored at many locations in a computer network, furthermore all the network subscribers should be able to access and update all this data. Many limited versions of these systems have already been realized, however no one has yet succeeded in designing a system that has all the characteristics of a single large data base. This requires the transfer of an extremely large amount of data between the various computers in the system which in turn necessitates powerful CPUs. Judging by the present development of public data networks it seems likely that data transfer costs will decrease sharply and that sufficiently high transmission rates will be obtained in the years to come.

Remarkable progress has been made in the data acquisition of information systems: Machines can now read typewritten text or text written in capitals at a much lower cost than before. This means that devices which are capable of processing written information more efficiently than ever before will soon be appearing on the market.

Software systems which can analyse natural language could form the basis for the automatic extraction of descriptors from documents, it should also be possible to construct dialog systems which can communicate with data bases in natural language.

FACTEURS NON TECHNIQUES QUI INFLUENCERONT LES SYSTEMES D'INFORMATION

par
A. YANEZ
Sous-Directeur Adjoint de l'Information Scientifique et Technique
CEDOCAR
26, Boulevard Victor
Paris Armées
75996
France

RESUME

On rappelle le contexte général de l'évolution vers l'ère de l'information. Tout converge pour montrer que le marché de l'information sera porteur, si nous savons construire de bons systèmes d'information et agir dans ce but sur les facteurs d'organisation, les facteurs politiques, les facteurs économiques et les facteurs juridiques. Ces quatre points ne sont pas traités exhaustivement, mais on s'efforce de mettre en relief certaines actions considérées comme prioritaires, signaler des écueils à éviter, et préparer ainsi dès à présent les systèmes d'information des années 1980.

CONTEXTE GENERAL DE L'EVOLUTION

Souvenez-vous des années 60. A cette époque le système d'information était présenté comme un tout parfaitement autonome (Planche 1) : une population d'utilisateurs et un corpus censé couvrir les besoins de cette population. L'accent se portait sur les notions de rappel et de précision, avec les tests de Cranfield, chers à Cleverdon, abondamment discutés en 1965 au "Nato Advanced Study Group" de la Haye sur l'évaluation des systèmes d'information.

Les systèmes d'information des années 80 seront ceux qu'annonce le rapport Nora-Minc (1) dont je ne citerai que de courts extraits : "l'informatique de masse irriguera demain la société comme le fait l'électricité (...) A la différence de l'électricité, la télématique ne véhiculera pas un courant inerte mais de l'information, c'est-à-dire du pouvoir. Elle modifiera le comportement des organisations et de la société toute entière (...)."

On pourrait aussi rappeler la thèse de Porat (2) ou bien les études de Machlup et de l'OCDE qui montrent que le taux de croissance du secteur informationnel est et restera beaucoup plus élevé que celui des autres secteurs, qu'on le mesure en termes d'emploi ou de valeur ajoutée (Planche 2).

Les systèmes des années 90, auxquels nous allons nous préparer pendant la décennie 80, seront probablement ceux qu'annonce Lancaster (3) (Planche 3) lequel, s'appuyant sur une solide enquête de Delphes, nous prédit que le Scientifique aura un terminal à son domicile, un terminal sur son lieu de travail... et un terminal qu'il transportera dans ses déplacements, pour ne pas se couper de l'information puisqu'il aura à prendre des notes, à échanger de la correspondance, à publier, à consulter des banques de données primaires et secondaires, le tout par voie électronique.

En attendant, les systèmes des années 80 vont assurer la transition et seront du type décrit par la Planche 4, c'est-à-dire que des courtiers ou des cellules de professionnels de l'information vont s'interposer entre l'utilisateur et les sources d'information, tout le temps qu'il faudra a cet utilisateur pour apprendre à puiser lui-même dans les ressources qui lui sont offertes à profusion mais de façon plus ou moins adroite. Nous allons avoir à agir sur les facteurs d'organisation, politiques, économiques et juridiques.. l'ensemble des facteurs humains qui influenceront cette évolution : vaste programme qui nous attend, et dont nous allons évoquer les différentes facettes.

1. FACTEURS D'ORGANISATION

UNISIST

Il s'agit tout d'abord de ne pas perdre de vue le programme UNISIST (4). Le système mondial d'information scientifique étudié et publié par l'UNESCO en 1971 énumère des objectifs. Les principes et recommandations qu'on y trouve n'ont pas vieilli dix ans plus tard, en dépit de l'évolution technologique, ce qui est plutôt rassurant. Le concept NATIS* permet de trouver dans chaque nation l'interlocuteur capable d'agir sur les facteurs d'organisation, dans un cadre coordonné, tout en tenant compte des priorités spécifiques.

---

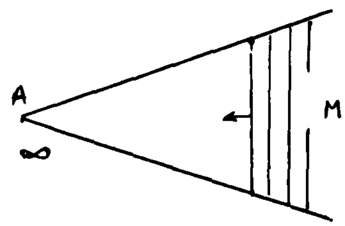
*NATIS : National Information System

SENSIBILISATION POUR UNE RESTRUCTURATION

Mais l'organe national de coordination (par exemple la MIDIST* en France) s'épuisera vite si son action ne suscite que l'indifférence et ne peut pénétrer la couche des cadres supérieurs. Auerbach et Slamecka (5) soulignent qu'aux Etats-Unis, il faut provoquer au plus haut niveau des Entreprises et de l'Administration, une prise de conscience de la signification et de la valeur de l'information et du potentiel offert par les systèmes d'information pour la stratégie d'entreprise et la conduite des échanges internationaux. Il faut donc alerter les dirigeants et mettre en place un programme et des moyens de sensibilisation : organiser pour les cadres supérieurs de l'ordre administratif et technique des séminaires, visites, exposés, conférences, démonstrations... et une formation montrant la valeur de l'information, son organisation et les moyens d'y accéder. Il faut arriver à une restructuration des entreprises qui prenne en compte l'information au niveau voulu et créer le réseau hiérarchisé d'information dans l'entreprise.

SYSTEME

Il faut aussi rappeler à ces cadres que diriger, piloter ou gouverner, c'est faire de la cybernétique et n'oublions pas que la cybernétique a été définie tantôt comme la science de l'information, tantôt comme l'art de rendre l'action efficace, et que tout système n'évolue bien que si l'on sait canaliser la rétroaction. Ceci est important parce que la cybernétique continuera à empiéter sur l'humain, et il n'y a pas lieu de s'en effrayer. En effet, à partir de la représentation que propose Aurel David (6) on peut imaginer M (les facteurs non techniques, ou l'outil, ou la machine) progressant vers A (Animus : l'esprit... les facteurs non techniques) pour prendre en charge progressivement des fonctions nouvelles de pilotage, sans danger pour A considéré comme le point de rencontre de 2 droites se coupant à l'infini. A gardant ainsi toujours sa signification transcendantale par rapport à l'outil.



Il faut donc savoir analyser le système, en le découpant en autant de parcelles qu'il est possible, comme le veut Descartes, mais sans perdre de vue que le tout est supérieur à la somme des parties, comme le rappellent Aristote ou Platon. Je proposerais que soit gravée partout la définition du système dont les implications sont fondamentales pour une saine évolution : "ensemble de conception homogène auquel est attribuée une mission précise dont l'utilisateur doit parfaitement connaître les possibilités et l'emploi optimal, conçu comme un modèle perfectible, avec recherche, à chaque instant, du meilleur résultat, par une succession de choix et de compromis dictés par des rapports de coût/efficacité en tenant compte des contraintes de tous ordres".

DECENTRALISATION

La nécessité d'une vision globale n'implique pas la centralisation. Il ne faut pas que des organisations se sclérosent et se fossilisent dans leur "Quête Mythique du Grand Système de Gestion Intégrée", sans voir que les logiciels complexes et rigides ne peuvent être que morts-nés, face au redéploiement constant des organisations humaines en environnement changeant. Alain Madec (7) cite cet ingénieur attardé qui, désireux de contrarier l'évolution informatique vers la décentralisation des fonctions disait : "Si Dieu avait voulu que l'intelligence soit distribuée, il nous aurait mis des cerveaux aux poignets...". Dans les années 80 se poursuivra cette décentralisation informatique faisant appel aux terminaux intelligents, à la mini-informatique, et à l'informatique répartie.

---

*MIDIST : Mission Interministérielle de l'Information Scientifique et Technique

SEPARATION DES FONCTIONS

Par opposition au système des années 60, on observe déjà en 81 une nécessaire différenciation des fonctions, comme ce fut le cas dans l'évolution du règne animal.

- Fonction Etudes (Besoins, Marché, Promotion, Prospective).
- Fonction Producteur de documents primaires.
- Fonction Producteur de banques de données textuelles ou factuelles.
- Fonction Serveur (regroupement et mise à disposition des banques de données)
- Fonction Courtier (recensement des sources - accès à ces sources - formation des utilisateurs).
- Fonction Formation.
- Fonction Pilotage (utilisation de la rétroaction pour agir sur les objectifs).

Fonction Etudes

Il est fondamental de se situer vis-à-vis des utilisateurs potentiels, d'avoir une politique commerciale en fonction de l'étude du marché, et de conquérir cette clientèle tant par l'efficacité (délais-compétence) que par une avance sur les concurrents (promotion-prospective).

Fonction Producteur de documents primaires

Auteurs et éditeurs doivent être rémunérés et fortement encouragés. Il ne faut pas créer de structure particulière pour produire l'information nouvelle, mais au contraire utiliser et encourager les structures existantes : Centres Techniques, Associations ou Fédérations d'Associations d'Ingénieurs, Laboratoires universitaires... Ce sont les milieux qui secrètent normalement l'information, s'ils sont aidés, encouragés, orientés et leur production canalisée. La production de documents primaires est le point sur lequel doit porter l'effort de chaque gouvernement puisqu'il s'agit du contenu des systèmes à construire, qui sera bon ou médiocre selon que le scientifique sera sollicité, encouragé, aidé, dans le cadre approprié, à produire thèses, rapports, articles, livres, brevets, mémoires de présentation dans les Congrès... ou qu'il sera incité à se taire et à rentrer dans sa coquille. L'adhésion et la participation active à une Association au moins devraient être considérées comme une nécessité pour chaque chercheur ou ingénieur. C'est le cas des Etats-Unis qui peuvent nous montrer des exemples à méditer : American Chemical Society - American Society for Metals, Institute of Electrical & Electronic Engineers, Engineers Joint Council... qui ont organisé leur système de publications primaires et secondaires, et le revendent dans le monde.

Fonctions Producteur de banques de données textuelles ou factuelles

Si l'on décide d'entrer sur le marché de l'information secondaire, on doit définir le secteur, rechercher des appuis (coopérants) pour couvrir ce secteur, normaliser le format et l'information, imaginer des sous produits (magnétiques ou papier) et définir des objectifs de clientèle (serveurs pour les produits magnétiques, éditeurs pour les produits papier). Là encore, les Centres Techniques et Associations d'Ingénieurs sont probablement les points d'ancrage d'une action efficace et durable.

L'idéal est que le fichier ou la banque de données soit produit par ceux qui en seront les principaux utilisateurs et qui sauront le mieux orienter son contenu et assurer la formation des autres utilisateurs.

Fonction Serveur

Il importe que les banques de données textuelles ainsi produites soient placées et regroupées le mieux possible sur des serveurs peu nombreux. L'utilisateur sera ravi de pouvoir trouver son information sur un seul site au lieu d'avoir à faire le tour de tous les fichiers de tous les serveurs. L'effet de synergie est fondamental.

Pour les banques de données factuelles au contraire, on peut admettre qu'elles restent sur le site de leur création et de leur mise à jour puisque, dans la plupart des cas, les problèmes posés devront être étudiés par les spécialistes chargés de l'entretien de la banque, pour conduire à des stratégies de recherche et à des réponses analysées et interprétées de façon satisfaisante.

Fonction Courtier

Ce n'est pas encore dans les années 80 que les utilisateurs traiteront eux-mêmes tous leurs problèmes d'information. Pendant cette période de transition, les courtiers vont même se multiplier. Entendons par là les organismes qui assurent l'interface entre l'utilisateur final et les sources d'information. Ce peut être le service d'information interne à l'entreprise, qui s'est connecté aux sources accessibles dont il connaît les fichiers, les procédures et les difficultés d'accès... ou un Organisme extérieur à l'entreprise qui se charge d'apporter la réponse si on lui explique le problème (exemple NERAC*). En attendant que "l'utilisateur final" soit tout à fait adulte et compétent, il y a encore de beaux jours pour les courtiers.

Fonction Formation

Cette fonction ne peut pas continuer à être confiée seulement aux fournisseurs de banques de données ou aux courtiers. Elle doit être prise en main par les Gouvernements suivant un programme à établir.

Fonction Pilotage

C'est à partir des réactions des usagers, encore une fois que seront prises les décisions d'acheter telle ou telle banque de données, ou seulement d'y accéder, directement ou par l'intermédiaire de tel ou tel courtier. L'ennui si l'on délègue la fonction recherche d'information à un courtier c'est qu'on court le risque de ne jamais impliquer l'utilisateur final et canaliser sa réaction, alors que si cette fonction est interne à l'entreprise on aura de bons résultats grâce au couple utilisateur final et documentaliste... et peu à peu si des programmes de formation sont mis en place, si les logiciels permettent des accès plus faciles, et si une nouvelle génération d'ingénieurs, déjà sensibilisée en école et par l'emploi des systèmes d'information grand public émerge, il n'est pas impossible que les courtiers disparaissent et que l'on arrive progressivement au système envisagé par Lancaster... Pour assurer la fonction Pilotage, des regroupements seront encouragés : Industrie de l'information, Groupe de producteurs de banques de données, Association d'utilisateurs dont les diverses pressions rétabliront certains équilibres. Ceci est fondamental, puisque encore une fois, piloter c'est utiliser la rétroaction.

2. FACTEURS POLITIQUES

Une bonne part de l'action politique est déjà dictée par les recommandations UNISIST, et une politique s'élabore au niveau européen, ce qui permet d'éviter en Europe les initiatives isolées et les cloisonnements inutiles. Il appartient bien entendu à chaque Gouvernement d'apprécier ses priorités, mais en se plaçant dans le cadre tracé.

GOUVERNEMENT ET SECTEUR PRIVE

A chacun son rôle : le Gouvernement gouverne, c'est-à-dire fixe les objectifs en fonction des moyens, favorise ou subventionne les actions prioritaires, définit une politique de l'information et agit à ce titre sur les structures et les organisations pour faciliter l'exécution de cette politique, développe la recherche en science de l'information pour en tirer de nouveaux objectifs pour l'avenir.

Le secteur privé quant à lui, se voit confier la réalisation, l'exécution, la mise en place de systèmes viables, efficaces, générateurs de biens et notamment de profit. Le secteur privé est mieux placé pour assurer la gestion, sauf si un impératif s'y oppose (exemple du Secret de Défense). Aux Etats-Unis, ce principe est clairement exprimé par la directive A76 selon laquelle le Gouvernement fédéral s'appuiera sur le secteur privé pour publier et disséminer l'information scientifique et technique, sauf s'il est dans l'intérêt national d'agir autrement (8).

Le secteur privé tend à considérer l'information comme un bien économique, qui peut être distribué suivant les mécanismes du marché, comme tous les autres biens et services.

Les Gouvernements considèrent qu'il faut préserver, indépendamment des lois du marché, un accès équitable à l'information. Ceci n'est pas seulement un droit mais c'est nécessaire à la préservation d'une société libre.

On retombe ici sur le Freedom of Information Act qui a permis aux Etats-Unis de déplacer la frontière commerciale du marché et la frontière juridique du secret (17).

---

* New England Research Application Center.

On considère généralement que les instances internationales et les gouvernements ne se dessaisissent pas suffisamment d'activités d'information. Aitchison (9) remarque qu'aux Etats-Unis l'industrie de l'information est plus forte et plus diversifiée qu'en Europe. On pourrait en débattre, en notant toutefois que l'émergence de l'industrie de l'information a été un phénomène périphérique et une retombée de la solide organisation et prise en main de l'information aux Etats-Unis par le gouvernement (NASA - NSF - NTIS - etc...). L'industrie est en quelque sorte le revendeur, vers un marché plus large, national ou international, d'une information collectée à une autre fin. Par ailleurs, l'organisation de l'information est calquée sur l'organisation de la recherche et des différences notables existent sur ce plan entre Europe et Etats-Unis. Il est donc possible que des réorganisations du secteur recherche entrainent ipso facto une redistribution des activités d'information entre gouvernement et secteur privé.

COMPETITION

Il me semble donc qu'on ne peut pas parler de compétition entre gouvernement et secteur privé. On doit au contraire encourager la compétition et l'émulation entre organisations du secteur privé - ne pas favoriser les situations de monopole, ou de partage à l'amiable du marché, situations qui conduisent à la stagnation et à l'autosatisfaction. C'est probablement la raison pour laquelle les systèmes d'information ont si peu évolué au cours des dernières années : un produit relativement médiocre se vend. Pourquoi prendre des risques en le perfectionnant. C'est ce que j'ai cru comprendre du discours de clôture de Salton au dernier congrès de Cranfield, attitude un peu plus nuancée lorsqu'il dit "superficiellement, il peut sembler qu'il ne s'est rien produit dans les années récentes. Après tout le fichier inversé et la stratégie booléenne remontent à l'époque de la carte perforée et existent toujours dans la plupart des systèmes... mais une observation plus précise montre que les choses ont tout de même changé"... et il rappelle que de nouveaux algorithmes permettent d'aller au-delà (10) ce qui est vrai pour SMART, SPIRIT et d'autres logiciels permettant de balayer un texte avec classement de pertinence à la sortie et dans des conditions plus proches du langage naturel de celui qui interroge. En d'autres termes la compétition sur le terrain du logiciel doit se développer. Il faut donc probablement se réjouir de voir par exemple que la Commission Fédérale des Télécommunications (FCC) a déclaré que la concurrence était désormais libre sur le terrain de la télématique entre ATT (ou le Bell System) et IBM ; ATT avait un quasi monopole téléphonique mais ne pouvait pénétrer jusqu'ici sur un territoire qu'IBM considérait comme le sien, et ceci n'est qu'un exemple d'un vaste mouvement politique et de réorganisation qui tend à une redistribution des forces vis-à-vis du secteur de l'information et de ses supports, aux Etats-Unis comme en Europe.

NORMALISATION

Un facteur politique intervient nécessairement vis-à-vis de la normalisation. Je partagerais volontiers les règles d'or énoncées par Anderla (11) : la normalisation ne doit pas freiner la créativité, elle doit s'en tenir à l'essentiel et assurer ou permettre une certaine stabilité, tenir compte de toutes les parties intéressées et de leur avis, et permettre une évolution sans hiatus.

3. FACTEURS ECONOMIQUES

Nous savons que l'information est une denrée très spéciale, immatérielle, qui se valorise en fonction de son utilisation, au contraire des autres biens, et qu'on ne sait pas évaluer, de sorte qu'on évalue souvent son support. Le seul critère est de la rattacher et de la comparer au secteur auquel elle s'applique et de se demander : que peut-on consacrer à l'information scientifique vis-à-vis de l'enveloppe recherche, à l'information économique vis-à-vis de la compétition internationale, à l'information sur l'énergie vis-à-vis des problèmes d'énergie, etc... plutôt que de ramener la valeur de l'information au coût du traitement ou du support en oubliant que l'information est aujourd'hui le premier enjeu dans la compétition internationale.

Si l'on trouve que l'information coûte cher, on peut dégager des principes d'action qui en diminueront le coût ou la valoriseront davantage :

1. Rendre quasiment impossible la recherche conduite sans examen préalable de l'information déjà disponible sur le même sujet (par exemple pression NSF* pour pousser à l'utilisation de la banque SSIE en liant le financement de la recherche à une telle participation) et introduire dans les contrats de recherche une clause obligatoire de recherche bibliographique.

2. Créer en Europe des centres de redistribution de l'information du type NTIS. Le système SIGLE** pour la littérature grise est un exemple à suivre...

---

* National Science Foundation.

** System for Information on Grey Literature in Europe

3. Eviter le contresens qui conduirait à tout sacrifier à l'informatique en négligeant la finalité qu'est le flux de données. La technique n'est que le contenant. Nous allons vers la construction d'un nouvel équilibre mondial où les échanges d'information vont jouer un rôle central dans le monde où l'interdépendance des Etats ne cesse de croître. Les facteurs en présence sont d'une part la libre circulation de l'information (au nom de la liberté d'expression et du libre échange) et d'autre part la responsabilité des états souverains recherchant un équilibre résultant d'échanges mutuellement profitables.

MARCHE

Appleyard rappelle qu'on trouvait aux Etats-Unis 14 000 terminaux en 1969 : on en trouve 2 millions aujourd'hui et on en prévoit 10 millions en 1989, mais il y a déjà aux Etats-Unis 130 millions de postes de télévision qui feront partie du réseau intéractif de demain. C'est donc un changement d'échelle considérable pour l'industrie et pour l'emploi.

Un autre facteur d'expansion est le fait que la production documentaire augmente, dans les domaines non techniques ou scientifiques et notamment l'économie, ainsi que les besoins d'information, et couvre rapidement tous les domaines de l'activité humaine.

L'avenir offre donc des opportunités considérables tant pour les banques de données que pour les logiciels et les matériels puisque la notion d'architecture de réseau tend à obliger à considérer ces aspects comme un tout. Sur l'ensemble de ces points, il importe d'éviter une domination d'une entité politique par une autre, ce qui serait dangereux*.

POLITIQUE DE PRIX

La politique de prix est liée à la politique de produits. Il faut cesser de subventionner les produits dont le coût est de plus en plus élevé et notamment les publications sur papier**.

Il faut aussi faire perdre l'idée que l'information peut être gratuite, et cesser d'offrir des services gratuits, en faussant ainsi tous les exercices de comptabilité.

Il faut tirer parti des nouvelles technologies pour abaisser les coûts. Il faut aussi développer une politique active de sous-produits. L'effort se porte actuellement sur la recherche bibliographique en ligne mais il existe bien d'autres applications qui peuvent tirer parti du fonds constitué : recherches bibliométriques, veille scientifique et observations statistiques ou analyses de toutes sortes, facilités par la mini-informatique.

Ce qui importe dans cette politique de prix, c'est que soient toujours incluses dans le prix les redevances à verser aux auteurs et éditeurs, et que la marge de profit reste suffisante pour éviter le recours à la subvention. Ceci est possible compte tenu de l'élargissement prévisible du marché, des regroupements qui vont s'effectuer, de l'abaissement des prix des équipements et des mémoires, de l'accélération des débits des réseaux. Dans la compétition, les meilleurs doivent survivre, et il n'est pas nécessaire qu'ils soient très nombreux. Les serveurs ne sont d'ailleurs pas inquiets lorsqu'ils analysent la situation (12).

Enfin il faut que les gouvernements ne viennent pas faire une concurrence au secteur privé de façon abusive, c'est-à-dire hors des cas où il intervient pour des raisons d'équité pour rétablir un équilibre.

Le sujet est vaste, et mériterait bien d'autres développements, puisque tout montre qu'on évoluera vers une économie informationnelle, suivant l'analyse proposée par ESKL, LION et POGOREL (13) dont je cite un court extrait "Les grandes manoeuvres... autour des nouveaux medias mettent en évidence la recherche de nouveaux produits et de nouveaux modèles de consommation susceptibles de prendre le relais du modèle logement - électro-ménager - automobile, qui avait été le moteur de croissance des années 1945-1973. L'information apparaît ainsi comme le champ nouveau de la croissance au moment où les sociétés industrielles se heurtent aux limites posées par la matérialité des produits".

---

* L'Europe face aux technologies de l'information. Le Dossier de l'Europe. E 1980 - 3 mars 1980. Commission des Communautés Européennes.

** Les principaux éditeurs se sont d'ailleurs déjà rapprochés du secteur informatique, ou bureautique ou télématique, qui représente leur seul avenir, si l'on en croit Lancaster (3).

4. FACTEURS JURIDIQUES

Le problème majeur reste celui du copyright, mais d'autres problèmes se poseront, qui influenceront les systèmes d'information : flux transfrontières, responsabilité du fournisseur de données.

COPYRIGHT

Si l'information est encore trop rarement offerte sur le marché c'est parce que, comme le souligne MADEC (14), faute d'une protection suffisante de la propriété intellectuelle, la rétention est considérée comme la meilleure garantie contre le "piratage".

Contre la retention, il y a le "Freedom of Information Act", mais un véritable droit de l'information est à mettre en place pour couvrir tous les aspects juridiques posés notamment par l'évolution de la technologie qui permet tous les abus, toutes les manipulations et rend difficile le contrôle des flux de données. Il est possible qu'en cette matière, l'ordre international précède l'ordre interne et lui offre un canevas uniforme...

En l'absence de cette législation internationale, il convient d'adopter une attitude prudente et, en tout état de cause, d'appliquer la législation sur la propriété intellectuelle. Cette recommandation est importante si l'on veut qu'un certain laxisme qui s'instaure déjà ne conduise pas à des conflits. Par exemple Holmes (15) rapporte les propos de Martin HOWE, à l'Institute of Information Scientists le 29 septembre 1980, faisant valoir que "le copyright se trouve acquis à l'usager (vested with the user) qui a pris possession des données en vertu de la formulation complexe qu'il a eu à utiliser pour les trouver"... ce qui est fascinant selon l'expression de Holmes. Il serait urgent de trouver des solutions de compromis, des codes de déontologie qui pourraient être introduits dans les contrats ou protocoles entre usagers et serveurs, serveurs et producteurs de banques de données, producteurs de banques de données et éditeurs ou auteurs, s'il s'agit de personnes morales différentes. Toute une réglementation se développera au niveau des protocoles d'accès.

FLUX TRANSFRONTIERES

La conférence de Juin 80 tenue à Rome sur ce sujet faisait suite à la conférence SPIN qui recommandait l'examen :

- des initiatives nationales,
- des programmes internationaux de recherche sur les flux de données et l'harmonisation des réglementations nationales (16).

Les systèmes juridiques sont fondés sur la notion de territorialité. Les pays exerçant leur souveraineté en réglementant à l'intérieur de leurs frontières les activités tangibles ou visibles, mais l'information, transformée en unités binaires invisibles transmises sur de grandes distances va obliger à modifier les systèmes juridiques nationaux. Certains états risquent de dresser des "barrières invisibles aux échanges invisibles" parce que la libre circulation des données dépasse de très loin la notion de libre échange. Elle ouvre la voie, sans possibilité de contrôle, à la fraude, à l'espionnage, au dumping ou à l'évasion de bénéfices, peut entrainer des déséquilibres, une vulnérabilité, une altération de l'identité culturelle. Toutes ces idées sont développées par Madec (14). Ce contrôle ou cette circulation devrait s'exercer sans s'opposer aux cinq principes fondamentaux dont s'inspirent les Etats-Unis : 1) Libre circulation de l'information. 2) Protection de la vie privée. 3) Libre jeu des mécanismes du marché. 4) Libre échange. 5) Disponibilité d'installations et de services de télécommunications appropriés (17).

RESPONSABILITES EN MATIERE DE FOURNITURE DE DONNEES

La question de savoir dans quelle mesure le producteur de la banque de données est responsable de la qualité et de la fiabilité des données fournies s'est déjà posée et se posera à nouveau. Le serveur n'est certainement pas en cause. Les limites de responsabilité du producteur devront être définies dans le cadre d'une recherche qui aboutira a une codification et à un droit de l'information apportant sur ce point aussi la clarté souhaitable.

5. FACTEURS HUMAINS

En France on parle de la loi du moindre effort, et non pas de la loi de MOOERS pour expliquer qu'un système d'information tendra à ne pas être utilisé lorsqu'il sera plus pénible de l'utiliser que de s'en passer. Quoiqu'il en soit cette vérité est essentielle. Il faut donc en tenir compte si l'on veut que le marché de l'information se développe et attire l'utilisateur final.

On a déjà bien étudié les systèmes intéractifs et le comportement de l'utilisateur d'un terminal. Il est même regrettable que cet acquis n'ait pas été utilisé pour les systèmes VIDEOTEX qui ont refait leur apprentissage des mêmes erreurs.

L'ergonomie du système est essentielle. L'analyse conduit à isoler des défauts de taille, qui souvent peuvent être facilement corrigés : effet de pression, effet de vitrine, difficultés diverses de lecture de l'écran, d'utilisation du clavier, phobies ou allergies possibles...

L'ingénieur est aujourd'hui généralement angoissé à l'idée d'utiliser un clavier et de converser avec un ordinateur. Mais il faut bien dire que rien n'est fait pour le mettre à l'aise (messages abrupts du genre "invalid statement", "erreur de syntaxe". L'utilisateur est finalement absorbé par la "mécanique de l'interrogation" au point qu'il en oublie son sujet. Les délais entre les messages, les répétitions de procédures... tout cela peut entrainer la fatigue, l'ennui et la frustration définitive. Les années 80 devraient permettre d'arriver à une amélioration sensible de la jonglerie demandée à l'utilisateur qui devrait pouvoir explorer un fonds documentaire et être satisfait de l'efficacité, au regard de l'effort accompli. L'accès au document primaire restera le point noir tant que chaque serveur n'aura pas introduit la commande au terminal, et que chaque pays n'aura pas organisé l'accès au document primaire.

L'utilisateur voudra aller plus loin et constituer lui-même sa documentation en "siphonant" l'information issue de diverses sources. Des expériences en cours telles que SHOEBOX et AUTONOTE (18) permettent déjà d'étudier les comportements et d'en déduire des solutions appropriées offertes pour améliorer la communication entre l'homme et la source d'information.

L'objectif n'est en effet pas de plier l'utilisateur au système construit sans lui, ni de le contraindre à utiliser un équipement imposé qui perturbe son environnement. L'objectif est un système construit avec l'utilisateur, et évoluant avec lui et pour lui...

BARRIERES LINGUISTIQUES

Il faudrait revenir à l'UNISIST qui souhaitait, dans la recommandation 4 du Groupe I (Outils de l'intercommunication entre systèmes) que soient développés des outils qui servent au contrôle et à la conversion des langues naturelles dans la science et la technologie : normaliser les nomenclatures scientifiques sur une base internationale, mettre au point des vocabulaires canoniques ou thesaurus destinés aux systèmes d'information et faire progresser l'indexation automatique des documents.

Au lieu de s'en tenir à ces objectifs prudents permettant le balayage de la littérature secondaire et l'identification de son contenu, on a préféré poursuivre la chimère de la traduction automatique des documents primaires, avec des résultats variables, bons ou mauvais selon qu'on y croit ou qu'on n'y croit pas (il y a un lien entre la traduction automatique et la médecine non orthodoxe).

L'objectif limité que proposait l'UNISIST avait l'avantage de pouvoir "accrocher" chaque langue nouvelle à un système relativement simple de normalisation des concepts sans entrer dans l'analyse de chaque langue (TITUS est un excellent exemple de cette approche prudente).

Je continue de penser que, malgré les progrès indéniables réalisés dans la traduction automatique, l'objectif limité de l'UNISIST permettrait déjà une certaine intercommunication des systèmes, suffisante pour se tenir au courant de la production de chaque groupe linguistique, et décider alors de traduire éventuellement tel ou tel document primaire... En matière de traduction automatique, le premier objectif devrait être de balayer la littérature secondaire, en réglant par là aussi le problème des coûts élevés de saisie puisque l'information dans ce cas existe déjà sur support lisible par calculateur.

CONCLUSION

J'ai essayé d'évoquer les facteurs qui influenceront les systèmes d'information de la décennie dans laquelle nous venons d'entrer.

L'enjeu est considérable et le marché est en pleine expansion.

Il nous appartient d'agir dès à présent pour que ces facteurs aient des influences favorables.

Trop d'aspects ont été évoqués ici en trop peu de temps. Des analyses beaucoup plus approfondies ont déjà été faites, notamment par les auteurs que j'ai cités.

Si, après cet exposé, on me demandait de résumer en quelques mots les points sur lesquels l'effort principal doit porter, je répondrais sans hésitation :

1) d'une part faciliter l'accès aux banques de données en simplifiant les procédures, en introduisant le langage naturel avec des logiciels beaucoup plus performants que ceux qui sont répandus aujourd'hui et en s'intéressant à d'autres familles linguistiques.

2) D'autre part, montrer aux ingénieurs et techniciens que l'information leur est indispensable et que l'accès à l'information est facile, et entreprendre une vaste campagne de sensibilisation et de formation, notamment au niveau des Ecoles et de l'Université, pour que, dans la compétition internationale déjà amorcée, les richesses aujourd'hui accessibles ne restent pas ignorées ou sous-utilisées, avec les périls que cela implique. Il ne s'agit pas seulement de formation mais d'une éducation devant conduire à une réelle prise de conscience de l'enjeu international, et conduisant à transformer de façon radicale le comportement des scientifiques des années 80 vis-à-vis de l'information.

Si ces objectifs sont poursuivis et atteints, le marché prendra soin de lui-même et la promotion des systèmes ne sera plus qu'une affaire d'émulation entre concurrents visant la meilleure satisfaction des besoins.

REFERENCES

(1) NORA (S.),MINC (A.). L'informatisation de la Société. Rapport à Monsieur le Président de la République. La Documentation Française ISBN 2-11-000-1577 et Annexes ISBN 2-11-000-161.5.

(2) PORAT (M.V.). The Information Sector: Definition and Measurement - Paper presented at the annual meeting of the ASIS - 1978

(3) LANCASTER (F.W. et Al). The Impact of a Paperless Society on the Research Library of the Future. Rpt. PB80-204548

(4) UNISIST. Abrégé de l'étude sur la réalisation d'un système mondial d'information scientifique. UNESCO - Paris 1971

(5) AUERBACH (I.L.), SLAMECKA(V.). Needed: Executive Awareness of Information Resources Information and Management 2 (1979) 3-6. North Holland Publishing Co.

(6) DAVID (A.). La cybernétique et l'humain. Collection Idées Gallimard -1965

(7) MADEC (A.J.). Objectifs et points d'appui d'une politique française en matière de flux transfrontières de données. Problèmes politiques et sociaux - ISSB 0015-9743 no 406 - Janvier 1981. La Documentation Française - Paris

(8) GANZ (C.). NSF Report of the Working Group on Private Sector/Government Relationships for Scientific and Technical Information - NSF/INT - 80101 June 1980

(9) AITCHISON (T.M.). An Analysis of the Present European Situation. ASLIB Proceedings 32 (Supplement) - January 1980

(10) SALTON (G.). Automatic Information Retrieval. Computer Vol. 13 No 9 - September 1980.

(11) Videotex in Europe. ISBN 0.904-933.22-9 Learned Inf. Ltd. Oxford

(12) CUADRA (C.A.). Surviving the Eighties: New Roles for Publishers, Information Service Organizations and Users. Paper presented 5 March 1980 at the 22nd NFAIS Conference

(13) ESKL (R.), LION (J.),POGOREL (G.). Essai d'analyse économique - Vers une économie informationnelle? Problèmes politiques et sociaux - no 406 - Janvier 1981. La Documentation Française - Paris

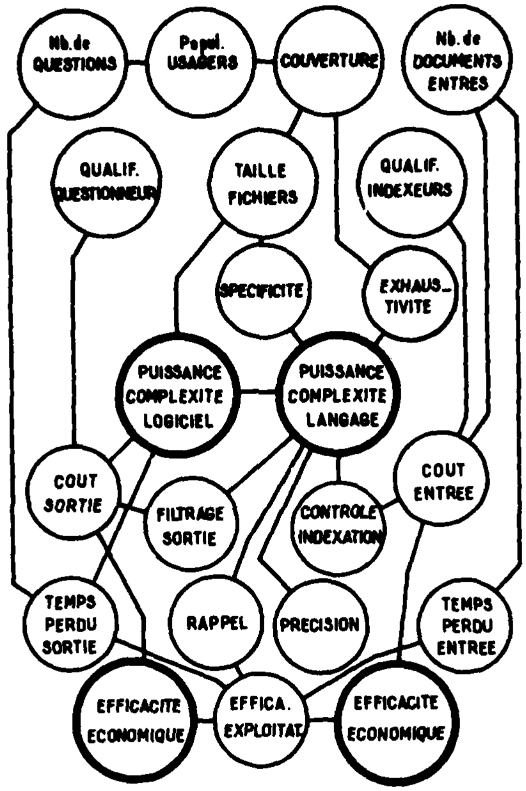
(14) MADEC (A.J.). Aspects économiques et juridiques des flux transfrontières de données. Problèmes politiques et sociaux, no 406 - Janvier 1981. La Documentation Française - Paris

(15) HOLMES (P.L.). The Re-use of Machine Readable Data and Copyright - a Pragmatic Approach to the Problems. Fourth International On-line Information Meeting - London 9-11 December 1980 - ISBN 0-904933-28.8 - Learned Information Oxford

(16) Bulletin de l'UNISIST - Vol 8 - no 4/80 - UNESCO Paris

(17) BUSHKIN (A.A.), YUROW (J.H.). Les fondements de la politique de l'information des Etats-Unis. Conférence de l'OCDE sur les politiques de l'information, de l'informatique et des télécommunications au course des années 80 - Paris, 6-8 octobre 1980

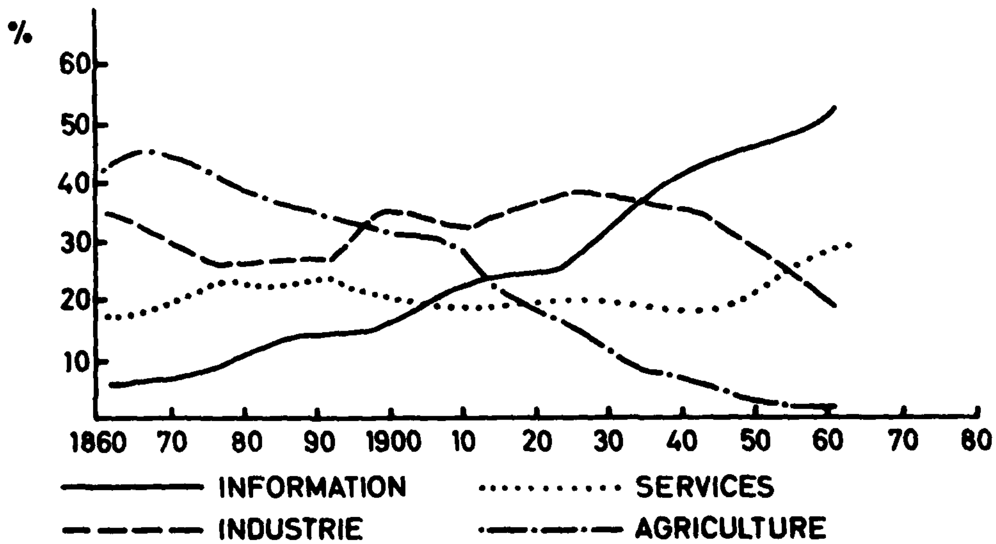
(18) LANCASTER (F.W.), FAYER (E.G.). Information Retrieval On-line. Wiley-Becker & Hayes Series Book. ISBN 0-471-51235-4.



<u>Planche 1.</u> Paramètres du système. d'après Climenson La Haye 1965

ERIC ED 122758
MARC PORAT 23.2.76

L'ERE DE L'INFORMATION



<u>Planche 2.</u> Répartition en 4 secteurs de la main d'œuvre Américaine (1860-1980)

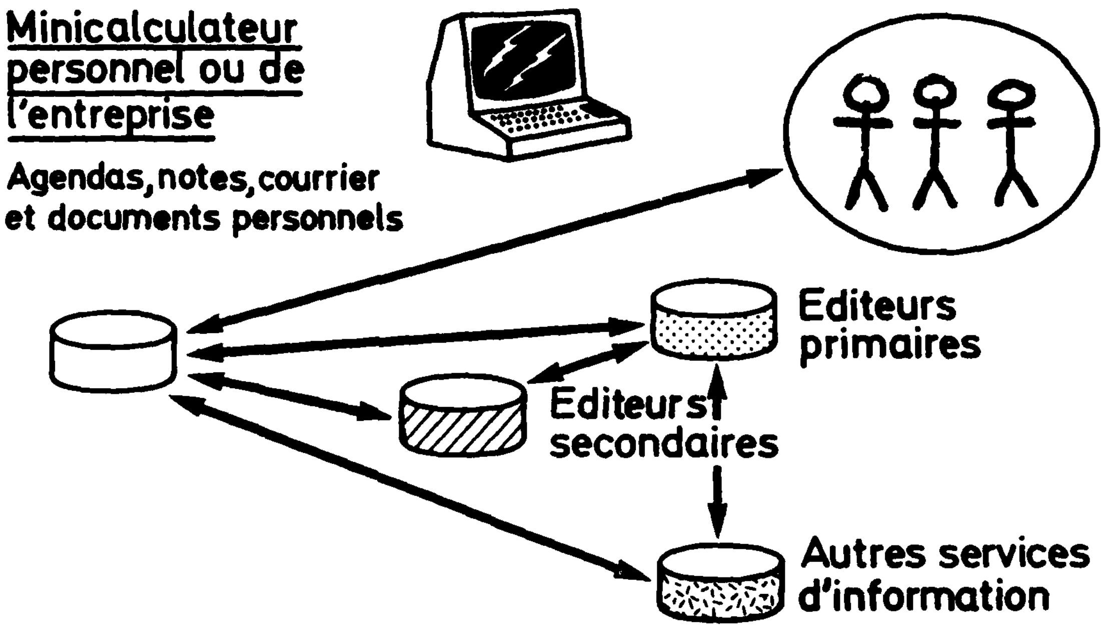


Planche 3. Systèmes d'information. Années 90
d'après F.W. Lancaster

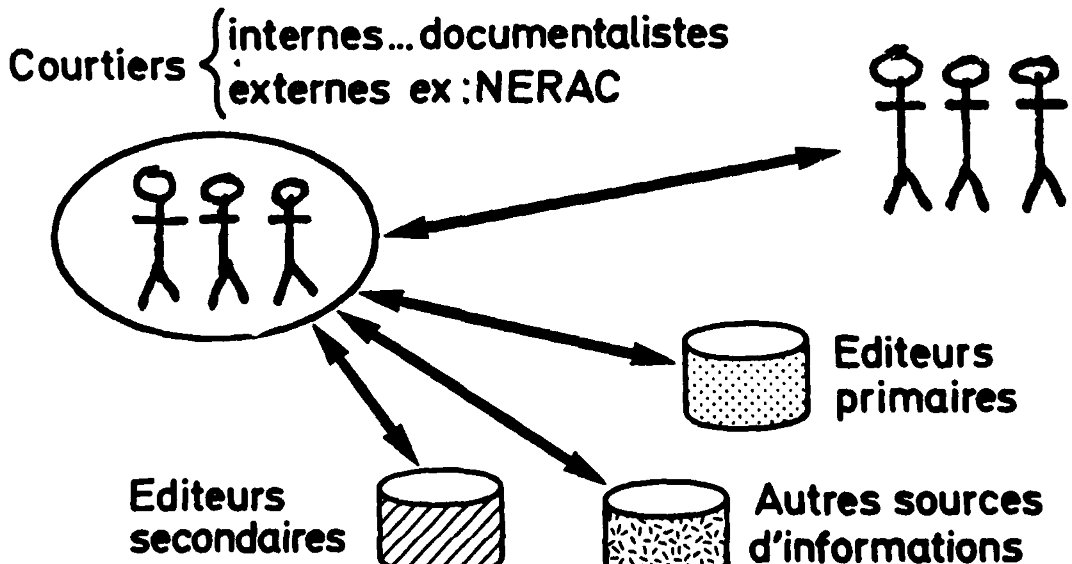


Planche 4. Systèmes d information. Années 80

NON-TECHNICAL FACTORS INFLUENCING INFORMATION SYSTEMS

by

M. Albert Yanez

Joint Deputy Director of Scientific and Technical Information
CEDOCAR
26 Boulevard Victor
75996 Paris Armées
France

(This English-text version of the Paper presented by M. Yanez has been produced in response to requests made by a number of persons who attended the Munich Symposium. It is regretted that time did not allow English-text versions of the illustrations to be produced. Accordingly, readers are referred to the French-text figures which appear on the immediately preceding pages.)

SUMMARY

A look backwards to the 60's and a reminder of the evolution which led to the so-called information era will help in setting the general framework. The information market of the 80's will obviously be buoyant if we are able to go on building good information systems and if we act accordingly on the organizational, political, economic, legal and human factors involved. These themes are not dealt with exhaustively, but the purpose is to highlight those actions which are felt to be essential. A number of pitfalls will be avoided if we do not put off the sound preparation of the systems of the 80's.

GENERAL FRAMEWORK OF THE EVOLUTION

Can you remember the 60's? In those days, the information system was looked upon as a self-contained entity (Fig.1): a "user population" and a "corpus of abstracts" supposed to cover the needs of the population. The accent was put on recall and precision, with the Cranfield Tests, endeared to Cleverdon, so thoroughly discussed at the 1965 NATO Advanced Study Course in The Hague, The Netherlands - a meeting dealing with the evaluation of information systems.

The information systems of the 80's are going to be those heralded by the Nora-Minc report (1) from which this short extract is taken: "mass informatics will permeate tomorrow's society just like electricity .... Unlike electricity, telematics will not carry an inert current but information, i.e., power. It will transform the behaviour of organizations and of society as a whole ..."

We could just as well quote Porat's thesis (2) or Machlup and OECD reports showing that the growth rate of the information sector is and will remain higher than that of other sectors, whether we look at it in terms of employment or of added value (Fig.2).

The systems of the 90's, which we are going to prepare in the 80's, will presumably be those envisioned by Lancaster (3) (Fig.3) who uses a sound Delphi Survey to predict that the scientist will have a terminal at home, a terminal in his office or lab ... and a terminal he will carry out on his journeys, in order to keep in touch with his notebooks, his mailbox, publishers, and data banks through an electronic link.

In the meantime, the systems of the 80's will be transitional, and of the type illustrated in Fig.4, which shows that brokers or other information professionals are going to interface between the user and information sources, until the time comes when the end-user has learned how to go it alone and to use the resources which are so profusely (and yet, up to now, awkwardly) offered to him. That's why we should be prepared to act on the various factors which are going to influence this evolution: it is a wide ranging programme, and we'll tackle some of its most obvious facets.

1. ORGANIZATIONAL FACTORS

UNISIST

First and foremost, we should not lose sight of the UNISIST programme (4). The world system of scientific information, as designed and published by UNESCO in 1971, lists a number of objectives. The principles and guidelines put forward in '71 are in no way obsolete 10 years later, in spite of technological changes, which is rather reassuring. The NATIS* concept makes it possible to find in every nation the counterpart formally entitled to influencing organizational factors, in a coordinated framework, and in the light of specific requirements.

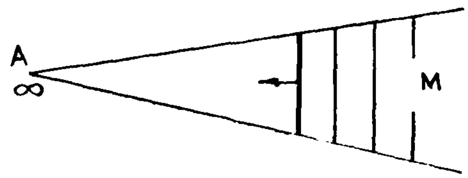
* *National Information System*

MANAGEMENT AWARENESS IN VIEW OF THE NEEDED REORGANIZATION

The national focal point, such as MIDIST* in France, will soon be neutralized if its efforts give rise to nothing but indifference, or cannot permeate the management stratum. Auerbach and Slamecka (5) stress that in the United States lack of management awareness remains a problem, so that many firms and public services are unable to assess the full meaning and value of information and the potential of information systems for the firm's strategy and the monitoring of international trade. Thus, we should alert managers through a programme aimed at them: it could be in the form of seminars, visits, lectures, exhibitions and demonstrations, and some sort of training on the subject of information, its organization, and the proper way to use it. This will lead to a reorganization of the firms, which will take care of information at every level and will tend to build a hierarchical network of information tailored to the organization.

SYSTEM

We should also remind management of the fact that to direct, steer, or govern is to step into cybernetics, which has been defined as the science of information, or the art of making action efficient. The system is healthy if the proper feedback is obtained. This is more and more important, as cybernetics will go on moving towards what had previously been human territory, and this should not frighten us. We can obviously use Aurel David's (6) representation M of the technical factors, or the tool, or the



machine moving towards A (animus or spirit or non-technical factors) to take care of new steering functions ... without any danger for A which is the point where two parallels intersect. A will no doubt maintain its transcendant value as opposed to the tool.

In the '80's, we must be able to analyze the system, and to cut it into as many parts as possible, as suggested by Descartes, but we must not lose sight of the fact that the whole is superior to the sum of the parts, as emphasized by Aristotle or Plato. I would suggest that a definition of the system and its fundamental implications be engraved wherever possible: "a homogeneously designed ensemble to which a specific mission is given, the possibilities of which are fully known by the user, who is well aware of its optimal use, from which the best result is looked for, at any point in time, through a succession of choices and trade-offs inspired by cost-efficiency relationships, with due consideration of all existing constraints".

DECENTRALIZATION

The requirement of a global vision does not imply centralization. Organizations should get rid of their sclerosis and the fossilized condition which derives from their "Mythical Quest for the Grand System of Integrated Management", without seeing that complex and rigid softwares can only be still-born, in view of the unending redeployment of human organizations in their moving environment. Alain Madec (7) makes references to this backward-looking engineer who was eager to thwart the normal trend of data-processing towards a decentralization of functions, and was heard saying "If God had wished intelligence to be distributed, he would have put brains in our fists ..." However, in the 80's, the decentralization and use of intelligent terminals, minicomputers and distributed data processing will have to be enforced.

SEPARATION OF FUNCTIONS

As opposed to the systems of the 60's, we are already experiencing a move towards the differentiation of functions, just as occurred in the animal kingdom:

- A design function (including users' needs, marketing, long-term studies)
- A primary document producer function
- A data base (textual or factual) producer function
- A host computer (or vendor) function
- A broker function (inventory of sources - best access to the appropriate sources - training of users)
- An education function
- A steering function (management based on systems analysis and use of feedback).

Design Function

It will remain fundamental to find one's proper place in relation to potential users, to have a marketing policy dictated by the market study, and to retain customers through efficiency as well as by maintaining an edge over competitors.

---

* *Joint-Ministerial Organization for Scientific and Technical Information*

Primary Document Producer Function

Authors and publishers must be rewarded and strong incentives must be found to enable this to be done. There is no need to create a new structure for this function. On the contrary, there is a strong case for using and reinforcing the existing structures: technical centres, Engineering and Scientific Societies and their federations, University laboratories and their scientists, these are the bodies which normally produce information, if they are aided, encouraged, oriented and if their material is collected. The production of primary documents is the item which should rank first in every government's efforts since obviously the contents should be there if a container is to be provided. The documents will be of high or mediocre quality according to whether the scientist is invited, encouraged and helped, in the appropriate framework, to produce theses, reports, articles, books, patents, papers to be discussed in Symposia ... or he is just ignored and ends up withdrawn into his shell.

An active participation in and membership of one scientific society at least should be a must for any engineer or scientist. Such is the case in the United States and a lot of pondering is needed in Europe over the achievements of such Societies as the Institute of Electrical and Electronics Engineers, the American Chemical Society, the American Society for Metals ... the Engineers Joint Council, whose primary and secondary publications are organized and redistributed or sold throughout the world.

The Database Producer Function

When you decide to enter the market of secondary information, you will have to define your sector, look for associates in a cooperative venture, provide full coverage of the sector, standardize your formats, look for by-products. There again, Technical Centres and Engineering Societies will probably be sound anchor-points for efficient and durable action. The aim is to have files or data banks produced by representatives of the potential users, i.e., people who are most likely to be knowledgeable as to the orientation of the contents and the training of users.

Vendor or Host Computer Function

Textual data banks thus produced should be placed on a restricted number of hosts. The user will be delighted to find information in one place instead of having to shop around the various files of a number of host computers. Synergy is fundamental.

Conversely, factual data banks are best placed at the location where they were set up and are updated since, in most cases, problems raised by the users will require analysis by specialists attached to the data bank, and it is they who are best able to choose the right search strategies and comment upon the results in a satisfactory way.

The Broker Function

It is unlikely that in the 80's users will be able to solve by themselves all their information problems. In this transitional period, new brokers are even going to emerge and grow. By brokers we mean anybody who interacts between the end-user and the information sources. It may be a service offered inside the organization, by someone who provides a link with the available sources and is fully knowledgeable on the command languages, the contents of files, protocols and specific difficulties of access. Or it may be some agency outside the organization that is supposed to produce the answer once the problem has been thoroughly discussed with its staff of analysts (such as NERAC*).

Until the end-user is fully grown up and educated, there will be thriving days for brokers.

The Education Function

This function is too important to be left only to the care of data-base producers, hosts and brokers. It will have to be grasped by governments and introduced into the syllabus of schools and universities.

The Steering Function

Again, it is the users' feedback which will lead to the decision either to buy this or that database, or simply to access a host, directly or through a broker. The trouble is that, if the search function is handed over to a broker, the end-user runs the risk of being forever kept out of the game and his feedback will not be available as it should be, whereas if the search function belongs to the organization, the coupling of the end-user with an analyst will stimulate and train the user.

Now, if and when education programmes are provided, if command languages and software are improved and easier access to the terminal is provided, and if a new generation of

* *New England Research Application Center*

engineers has been educated at school and at home through the use of videotex sytems and home computers, brokers will become extinct and Lancaster's vision of the future will prevail.

In the framework of this steering function, mergings must occur, of the sort we are already witnessing: Information Industry association, Data-base producers associations, and users' associations, and their pressure will insure the proper balance between the acting forces. Their feedback action will be fundamental for speeding up improvements.

2. POLITICAL FACTORS

A significant part of the political action is already suggested in the UNISIST guidelines and a policy is being built at a European level, which should prevent moves being taken in isolation, and unnecessary partitioning. It is the task of every government to assess specific priorities and to act in the direction of the guidelines.

PUBLIC VERSUS PRIVATE SECTOR

The roles should be clear-cut: governments have to define objectives in accordance with the available means, to promote and subsidize top priority moves, state what the information policy will be and act accordingly upon structures and organizations in order to bring about an implementation of the policies and develop research in information science from which new objectives will be set for the future.

The private sector will be entrusted in essence with the implementation tasks and the operation of facilities needed for viable and efficient systems, creating new goods and profit. The private sector is best equipped for management except if an overriding consideration is at stake (for example, Classified information). In the United States, this principle is clearly stated in the A76 Act, which provides that the Federal Government will rely on the private sector for the publication and dissemination of scientific and technical information, unless it is in the national interest to act otherwise(8).

The private sector tends to look upon information as an economic good, which can be distributed according to the market mechanisms, just like any other goods or services.

Governments maintain that an equitable access to information must be preserved, independently of market laws, not only because freedom of access to information is a citizen's right but because it is essential for the preservation of a free society.

We here are falling back on the Freedom of Information Act which made it possible in the United States to displace the commercial frontier of the market and the legal frontier of the Classified data (17).

It is generally felt that international bodies and governments are prone to keep their information activities to themselves. Aitchison (9) observes that, in the United States, the information industry is stronger and more diversified than in Europe. We could take issue on this, and point out that the growth of an information industry has been a peripheral phenomenon and a spin-off of a sound and stubborn government action (COSATI, NASA, NSF, NTIS ...). The information industry appears as a middle-man organization, aiming at a wider national or international market for information initially collected for another purpose. Furthermore, the information organization is a carbon copy of the research organization, and the conditions in Europe are substantially different from those in the U.S.

It is possible that reorganizations in the research activities will entail a new distribution of responsibilities between government and private sectors.

COMPETITION

It seems to me that there cannot be any competition between the government and the private sector, where monopolistic situations should be avoided, as well as any friendly sharing of the market, or any situation leading to stagnation and self-satisfaction. This is probably the reason why information systems have shown few improvements in recent years: a mediocre product has its customers, so why should we take any risks by searching for improvements? This is what I gathered also from the after-dinner lecture given by Salton at the 1979 Cranfield meeting. He is somewhat more subtle (with his tongue in his cheek) when he writes: "Superficially, it may seem that little has occurred in information retrieval in recent years. After all, inverted file organization and Boolean query statements go back to the punched card era and still prevail in most systems. A closer look, however, reveals that on-line systems have changed the established order..." and he reminds us that new algorithms would make it possible to go a step further (10), which is true with SMART, SPIRIT and other softwares which would make it possible to scan the literature and deliver a proper ranking of relevant documents at the output, starting the query in the plain language of the user. In other words, competition in the software field should develop. And perhaps we should rejoice over the Federal Communications Commission's (FCC) statement that competition is now open in the field of telematics between ATT (or the Bell System) and IBM. ATT had a

monopolistic position with regard to telephones but could not up to now penetrate a territory looked upon as IBM's. This is merely one example of a vast political and organizational drive tending to a redistribution of forces in the information sector which has started in the United States as well as in Europe.

STANDARDIZATION

Politics are necessarily involved in standardization activities. I would agree with the golden rules stated by Anderla (11): standardization should not be a hindrance to creativity, and should essentially ensure or permit a certain stability, take good account of the opinion of all the parties concerned, and enable a steady evolution.

3. ECONOMIC FACTORS

We all know that information is a very special, intangible commodity. Its value is related to its degree of use. Yet we often value the recording medium instead of the information it contains. The only criterion for assessing its value is to relate it to the sector which produces or uses it. We may ask ourselves what amount of money should be devoted to scientific and technical information in relation to the amount devoted to research, or to economic information in relation to the requirements of international economic competition, rather than any assessment related to the cost of processing, thus losing sight of the fact that information is much more than paper or tape. It is the prime commodity at stake in present and future international competition.

If people find that information is costly, I suggest we apply revised standards of judgement which will value it more and thus decrease its relative cost:

1. Let us make it impossible to start any 'new' piece of scientific research without a prior investigation of what has already been done and written on the subject (something like the National Science Foundation pressure exerted on private and government research organizations to urge them to use SSIE if they want to obtain any research grants).

Let us also enforce the introduction in research contracts of a standard clause which makes it mandatory to start research only when a bibliographic survey has been conducted.

2. Let us create in Europe centres, such as NTIS, for the redistribution of information. The SIGLE* System is a good example.

3. Let us avoid any misunderstanding which could lead to the sacrifice of information on the altar of informatics, so that we lose sight of the prime objective, namely, the flow of information. The technical aspect, the equipment, is nothing but the container. We are heading towards a new world balance in which information exchanges are going to play a central part and where the interdependence of states will increase. The existing factors are on the one hand the free circulation of information (in the name of liberty and free trade) and on the other the responsibility of sovereign states to seek a balance based on mutually profitable exchanges.

MARKET

Appleyard reminds us that there were 14,000 terminals in the United States in 1969. Today there are two million and some ten million are predicted for 1989. But there are already in the United States 130 million TV sets which will become part of the interactive network of tomorrow. There will thus be a considerable change of scale for industry and employment.

Another expansion factor is the fact that document output is increasing, in non-scientific sectors, and in particular in the economic sector, in tune with a general reliance of the individual upon information, which is thus spreading to all sectors of human activity.

The future offers huge opportunities, for data banks as well as for software and hardware, which cannot be dissociated since the very concept of network architecture requires consideration of the overall system. On all these points, no one political entity should be allowed to prevail over any other, as this would be dangerous.**

PRICING POLICY

Pricing policy is tied to products policy. It will become necessary to stop subsidizing products whose costs are constantly escalating. In particular, this applies to hard-copy publications.***

The belief that information should be free will disappear and we should from now on stop offering products or services free of charge, as this practice makes nonsense of all accountancy exercises.

---

* System for Information on Grey Literature in Europe

** L'Europe face aux technolgies de l'information. Le Dossier de l'Europe. E 1980 3 March 1980. The European Common Market Commission.

*** The principal publishers have, by the way, already become involved in the fields of informatics, telematics, etc., which, if Lancaster (3) is to be believed, represent their only future.

A decrease in prices will result from a better use of the new technologies, also from the search for more by-products. *Efforts are now centered on* the on-line retrospective search, but there are hosts of uses which can be made of a data bank, such as bibliometric studies, alert systems, statistical observations and analyses of all sorts, which are made easier by the use of mini or microcomputer systems.

What is important in arriving at pricing policies is that full consideration is always given to authors and publishers, and that the profit margin is such that the need for subsidies is avoided.

This is possible, in view of the foreseeable widening of the market and of mergers which are bound to occur, on top of a decrease in the prices of equipment and storage units, and the speeding up of data flow rates on networks. In that competition, survival will be for the fittest, and there is no need to have great numbers of them. We should note that the hosts or information vendors are not really anxious when they analyse the present situation (12).

Last but not least, government service should not come on the stage as mere competitors of the private sector in an obtrusive way, that is in situations other than those where government action is needed for equity or balancing purposes.

The subject is wide-ranging, and deserves ample discussion, since it is obvious that we are moving towards an information economy, according to the thesis put forward by Eskl, Lion and Pogorel (13), from which this extract is taken: "The great manoeuvres which are developing around the new media are evidence of a search for new products and new consumer models likely to constitute the next stage in the relay race, taking over from the 'housing/home appliances/personal car/model' which had been the growth force of the 1945-1973 era. *Information* thus appears as the new growth territory at a time when industrial societies are being confronted with the limits imposed by the materiality of products."

4. LEGAL FACTORS

The main problem remains that of copyright, but other problems will emerge and will influence information systems: transfrontier information flow, for example, and the responsibilities of the providers of data.

COPYRIGHT

If information is still too scarce on the market, it is because, as stressed by Madec (14), the protection of intellectual property is insufficient and retention is looked upon as the best guarantee against piracy.

Against retention, there is the Freedom of Information Act, but an actual Information Law remains to be drawn up to cover all the legal aspects arising, in particular, from the progress in technology which makes it tempting to trespass the law when handling data, and which makes it difficult to monitor the movements of data. It is likely that, in this matter, international law will act as the harbinger of internal law and will offer the needed framework.

In the absence of international legal guidelines, our attitude should be one of prudence and, for all purposes, follow the strict enforcement of the present law on intellectual property. This recommendation is important *if* we want to avoid more of the permissive practices which are already in evidence and which could well turn into disputes. For example, Holmes (15) refers to a statement made by Martin Howe at the Institute of Information Scientists on 29 September 1980, to the effect that "copyright might be vested with the user who has taken ownership of the data by virture of the complex search formulation he has used to retrieve them" which is fascinating, in Holmes' own words. It would be *urgent* to find trade-offs and codes of practice which could be inserted in contracts or protocols between users and vendors, vendors or hosts and data-base producers, data-base producers and publishers or authors, when the latter are not part of the same corporate entity. A set of rules will have to be developed for access to information through agreed memoranda of understanding.

TRANSFRONTIER INFORMATION FLOW

The conference held in Rome in June 1980 on this topic was in the wake of a SPIN conference which recommended an examination of:

- national initiatives
- national research programme on data flows and the harmonization of national regulations (16).

Legal systems are founded upon a concept of territoriality. Countries exert their sovereignty by enacting laws within their frontiers on the subject of tangible or visible activities, but information, when transformed into invisible binary units transmitted over great distances, is going to bring about a modification of the national legal systems. Some states might raise invisible barriers to invisible

exchanges because the free transfer of data is far removed from the concept of free trade. It opens the way to fraud, without any possibility of checking it, to espionage, to dumping practices, or to the evasion of profit, and can thus bring about conditions of unbalance, vulnerability, or an alternation in cultural identity. All these points are developed by Madec (14). This control of transfer should be exerted without encroaching upon the five fundamental principles which inspire the United States: 1) free circulation of information, 2) protection of privacy, 3) free action of market mechanisms, 4) free-trade, 5) availability of the appropriate telecommunication utilities and services (17).

RESPONSIBILITIES VESTED WITH THE SUPPLIER OF DATA

The problem of the extent of liability incurred by the data bank supplier as regards the quality and reliability of the data he provides has already been raised and is going to be discussed again. The host is certainly not involved. The scope or limits of responsibility of the data-base supplier should be the subject of research leading to a code of practice as part of an information law, in order to bring on this last point the desired clarification.

5. HUMAN FACTORS

In France, we speak of the law of least effort, and not of Mooers' Law, to explain that an information system will tend not to be used whenever it will require more effort to use it than to do without it. Be that as it may, this truth is essential. We have to keep it in mind if we want the information market to expand and to attract more end-users.

Interactive systems and the behavioural aspects of the user of these systems have already been exhaustively studied. It is a great pity that these observations have not been used for the Videotex Systems currently under review, which have gone through the same teething troubles and errors.

The ergonomic aspects of system use are essential. Sizeable defects have been found through observation and most of them can be corrected: pressure effect, fishbowl effect, peephole effect, legibility and readability of the information displayed on a screen or on a printer and a host of defects of this sort which give rise to allergies or phobias.

Today's engineers are most often panicked by the very thought of using a keyboard and of conversing with a computer. But we must recognize that little is being done to make them relax: stern messages appear such as "invalid statement" or "syntax error". The user is eventually engrossed by the mechanics of the "dialog" to the extent that he loses sight of his initial problem. The time lag between messages, the repetition of arid procedures. All this leads to weariness, dissatisfaction and eventual frustration.

The 80's should bring a significant improvement in the juggling capabilities systems expect from the end-user, who should be able to scan a file with satisfaction as to the efficiency-to-effort ratio. Access to the primary document will remain a matter for concern as long as the command for ordering documents is not available at the terminal - that is, as long as every country has not organized proper access to primary information.

The end-user is going to be asking for more, and will wish to be able to handle his own documentation in his own way by "siphoning" on his intelligent terminal information from various souces. Such experiments as Shoebox and Autonote (18) are under way and are used to assess the behaviour of users and infer solutions likely to improve communication between the scientist and his information sources.

The purpose, in effect, is not to make the user accept the shape of a system built in his absence, or to try, under duress, to lure him into using an equipment which disturbs his environment. The purpose is to let him suggest what the system should be and how it should grow.

LANGUAGE BARRIERS

Here we should come back to UNISIST and a wish expressed in guideline 4 of Group I recommendation (Tools for intercommunication between systems): Tools should be developed for the control and conversion of natural languages in science and technology, i.e., a standardization of scientific nomenclatures on an international basis, the development of canonic vocabularies or thesauri used in information systems, and the advancement of automatic indexing.

Instead of working within these limited objectives aimed at scanning secondary literature in order to identify its contents, the chimera of automatic translation was preferred, with variable results, good or bad according to whether you are a believer or not (there is an affinity between automatic translation and unorthodox medicine).

The limited objectives set by UNISIST represent common sense, in view of the fact that new languages could gradually have been entered in to this simple system (Titus is an excellent example of this prudent approach).

I am still convinced that, despite the obvious advance noticeable in automatic translation, the limited objectives set by UNISIST would permit a better intercommunication between systems of different languages, and would be enough to get an awareness of the production of each linguistic group .... with a view to translating what is most relevant. I suggest the first step remains the scanning of secondary literature, which also solves the problem of high input costs, since the information is available in machine-readable form.

CONCLUSIONS

I have been trying to point out what factors will be influential in the decade we have just entered.

The stake is tremendous and the market will go on expanding.

It is our duty to act now in order to make sure that these factors will have favourable influences.

Too many facets have been raised here in a short time.

Further consideration of these problems is needed, starting in particular from UNISIST and the authors I mentioned.

If you were to ask me what are, in a few words, the facets which require most attention, I would unquestionably answer:

1. first, ease access to data banks by simplifying procedures and by using natural language with more efficient softwares than today ... and be more attentive to other languages.

2. Remind scientists and engineers of the fact that information is their normal food, and undertake a campaign aimed at schools and universities in order to make sure that, in the current international competition, no information resources are ignored or under-used. This is more than a matter of training. It is a real education which should lead to the full awareness of the stake involved, and a resulting transformation of the behaviour of the scientists of the 80's with regard to information.

If these objectives are pursued and attained, the market will no doubt take care of itself and the marketing of information systems will simply be a matter of emulation between competitors eager to work for a better satisfaction of their users.

REFERENCES

(1) NORA (S.),MINC (A.). L'informatisation de la Société. Rapport à Monsieur le Président de la République. La Documentation Française ISBN 2-11-000-1577 et Annexes ISBN 2-11-000-161.5.

(2) PORAT (M.V.). The Information Sector: Definition and Measurement - Paper presented at the annual meeting of the ASIS - 1978

(3) LANCASTER (F.W. et Al). The Impact of a Paperless Society on the Research Library of the Future. Rpt. PB80-204548

(4) UNISIST. Abrégé de l'étude sur la réalisation d'un système mondial d'information scientifique. UNESCO - Paris 1971

(5) AUERBACH (I.L.), SLAMECKA(V.). Needed: Executive Awareness of Information Resources Information and Management 2 (1979) 3-6. North Holland Publishing Co.

(6) DAVID (A.). La cybernétique et l'humain. Collection Idées Gallimard -1965

(7) MADEC (A.J.). Objectifs et points d'appui d'une politique française en matière de flux transfrontières de données. Problèmes politiques et sociaux - ISSB 0015-9743 no 406 - Janvier 1981. La Documentation Française - Paris

(8) GANZ (C.). NSF Report of the Working Group on Private Sector/Government Relationships for Scientific and Technical Information - NSF/INT - 80101 June 1980

(9) AITCHISON (T.M.). An Analysis of the Present European Situation. ASLIB Proceedings 32 (Supplement) - January 1980

(10) SALTON (G.). Automatic Information Retrieval. Computer Vol. 13 No 9 - September 1980.

(11) Videotex in Europe. ISBN 0.904-933.22-9 Learned Inf. Ltd. Oxford

(12) CUADRA (C.A.). Surviving the Eighties: New Roles for Publishers, Information Service Organizations and Users. Paper presented 5 March 1980 at the 22nd NFAIS Conference

(13) ESKL (R.), LION (J.),POGOREL (G.). Essai d'analyse économique - Vers une économie informationnelle? Problèmes politiques et sociaux - no 406 - Janvier 1981. La Documentation Française - Paris

(14) MADEC (A.J.). Aspects économiques et juridiques des flux transfrontières de données. Problèmes politiques et sociaux, no 406 - Janvier 1981. La Documentation Française - Paris

(15) HOLMES (P.L.). The Re-use of Machine Readable Data and Copyright - a Pragmatic Approach to the Problems. Fourth International On-line Information Meeting - London 9-11 December 1980 - ISBN 0-904933-28.8 - Learned Information Oxford

(16) Bulletin de l'UNISIST - Vol 8 - no 4/80 - UNESCO Paris

(17) BUSHKIN (A.A.), YUROW (J.H.). Les fondements de la politique de l'information des Etats-Unis. Conférence de l'OCDE sur les politiques de l'information, de l'informatique et des télécommunications au course des années 80 - Paris, 6-8 octobre 1980

(18) LANCASTER (F.W.), FAYER (E.G.). Information Retrieval On-line. Wiley-Becker & Hayes Series Book. ISBN 0-471-51235-4.

AUTOMATED INDEXING AND THE FREE-TEXT METHOD

Dr Werner Meder
Gesellschaft für Information und Dokumentation mbH (GID)
Sektion für Technik (SfT)
Herriotstr. 5
D-6000 Frankfurt 71

In broad terms, automated indexing for document content analysis is a method intended to supersede or supplement the knowledge and experience of the human analyst. It requires:

- the construction of programmable rules for the analysis of texts, and
- the development and application of a relational dictionary connecting the words in the texts to the controlled indexing vocabularies.

The unresolved difficulties of constructing such large dictionaries are the main reasons why automated indexing has not as yet emerged from the laboratory.

Free-text search has been developed for application to documents not intellectually analyzed and as a supplement to search based upon intellectually indexed documents. Both the increasing costs and inconsistent quality of intellectual indexing and the unsatisfactory situation with regard to automated indexing has led to the increasing application of free-text search. However, the association between an initial query formulation and its free-text alternative terms is just the opposite to that of the indexing vocabulary of documents. Both methods - automated indexing and assistance to free-text search - are involved with similar problems: to substitute or reduce intellectual work using stored information in forms of relationships between words. Both techniques can be regarded as complementary and aimed towards achieving improved retrieval.

Introduction

About ten years ago, G Salton (1), along with other experts, expected that automated indexing would be routinely applied in the near future. As is now obvious, this expectation proved overly optimistic. The questions whether computerized content analysis is practically feasible for general applications and when a proven system for content analysis will be achieved remain very much open.

Today I should like to summarize what the problems of intellectual and automated indexing are. In addition, I intend to discuss why the expected development of computerized content analysis was delayed and what has to be done to transform existing experimental systems scattered throughout the world into standardized tools applicable to information retrieval. Details concerning the different approaches and problems in automated indexing are omitted; they are discussed elsewhere (e.g. (2), (3)).

Any description of the state of automated indexing must include a discussion of the free-text method for information retrieval. I hope to make clear that both methods - automated indexing and free-text retrieval - are essentially related as to their problems and methods.

The Concept of Content Analysis

The analysis of documents for the purpose of information retrieval includes two main steps:

1. the determination of what a document is about;
2. the translation of this conceptual analysis into an appropriate documentation language, independently of whether a classification system is used or controlled vocabularies as usually found in thesauri.

To accomplish the first step, the knowledge of an expert in the field covered by the documents is required. The second step requires an expert in the special field of documentation. Finally, both functions have to be integrated by one person - the indexer.

Thesauri indexing results in the intellectual assignment of a set of descriptors to each document. In most cases, no relationships are indicated between such thesaurus descriptors. The accomplishment of indexing good enough so that subsequent retrieval will achieve acceptably relevant search results can only be regarded as an art.

Over the last decade, three factors in document analysis have become increasingly important:

1. the immense increase of the production of scientific and technical literature;
2. the rapidly growing costs per document for indexing and classification work; and
3. the lack of qualified specialists capable of meeting the requirements stated above.

The existing bottleneck in documentation work involving content analysis is a direct consequence of the above factors.

During this same time period, printing technology was developed which uses machine-readable data for computer editing and subsequent photo-typesetting. As a side-product, machine-readable databases on magnetic tapes have become available consisting of abstract journals, current awareness services, and similar printed products. However, the material within these databases has either not been subjected to content analysis or has only been analyzed to the limited extent required for producing printed indexes. The intellectual analysis of such databases for information retrieval simply is not economically or humanly feasible.

The Problems of Automated Indexing

During the sixties, the lack of collections of machine-readable documents large enough to prevent misleading results was one of the main problems in conducting and evaluating experiments using automated indexing methods. Today this problem has been resolved. It now has become evident that quite a different quantitative problem with regard to automated indexing perhaps had been somewhat underestimated. Analogous to the situation with intellectual indexing, automated indexing requires an extensive set of instructions for accomplishing the conceptual analysis and for translating the results into the documentation language used. These instructions consist of:

- a list or "dictionary" of the complete free-text vocabulary of the subject field to which automated indexing will be applied; and
- a list of the semantic relationships between the free-text terms and those of the controlled vocabulary established within a consistent and sufficiently complete computer programmable model.

In addition, algorithms for the comparison between such lists and the document texts must be available. Other analytic features, e.g. for morphological and syntactical analyses, and procedures for the evaluation and assignment of indexing terms to documents must be provided.

Even this extremely simplified description of the requirements for automated indexing should indicate that the current bottleneck involves the strategies and methods for obtaining the needed instructions, and the integration of these into the dictionaries of the indexing system.

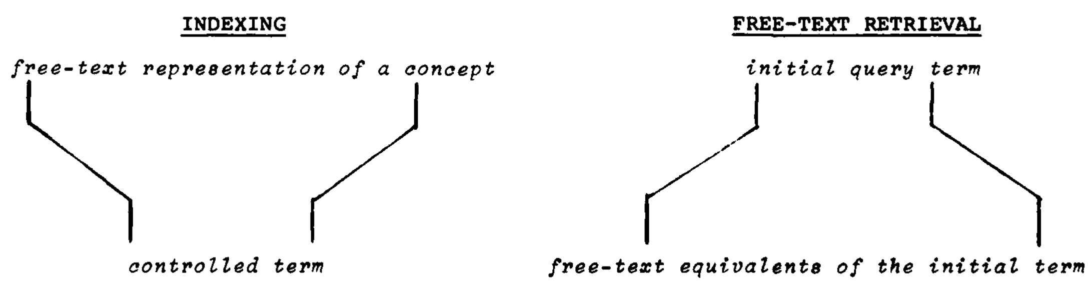
The Problems of Free-text Retrieval

By the early seventies it became evident, due to reasons in part just discussed above, that automated indexing would not replace intellectual indexing in the short-term future. At the same time, intellectual indexing proved too costly and, given the quantity of machine-readable documents, impractical except for certain special cases. The existing free-text retrieval systems became available as "emergency exits" from this nearly hopeless situation. However, one must now emphasixe that free-text retrieval essentially transforms all of the problems of content analysis into problems associated with search formulation and the retrieval process.

Free-text retrieval basically means that most or all parts of the stored documents, e.g. titles, abstracts, are searchable. Further, so-called context operators are available, functioning as surrogates for syntactical analysis, in addition to the well-known Boolean operators AND, OR, and NOT. Note that free-text systems offer the searcher no assistance in resolving semantical problems. As such, good retrieval - obtaining useful and relevant results - is dependent upon the experience, skill, intuition, and, of course, the good will of the searcher far more so than with descriptor retrieval.

Comparison of Both Methods

As a framework for some suggestions concerning future developments, a simplified comparison between the indexing and free-text retrieval concepts is offered.



Indexing is an a-priori method. The indexer has the opportunity of viewing the topics or subjects covered by the document in context. The full textual material of the documents is available to aid in selecting appropriate indexing terms. On the other hand, at the time of indexing the full range of potential aspects with which documents may later be retrieved from the perspectives of the variety of potential requestors of information typically is unknown or poorly stipulated.

In contrast, free-text searching is an in-situ method. For retrieval, the aspects for which the search is being conducted are known as formulated by the requestor of the information. As such, retrieval is independent of the subjective contributions of indexers assigning descriptors to documents which can at best only partially describe the full informational contents. With free-text search, it is the searcher left essentially on his own in selecting, evaluating, and determining the usage of terms with which a concept, topic or subject may be expressed or denoted in the textual material. In additon, the searcher has to deal with problems stemming from the varieties of syntactical structures found in textual material with only the primitive tools of context operators.

Suggestions Concerning Further Development of Automated Indexing

Earlier, it was observed that the fundamental problem for automated indexing today involves the development of methods for constructing dictionaries of words with their relationships established to the vocabulary of a given subject field. The task is to accomplish the above in a reasonable time with the assurance the the result is sufficiently complete, consistent, and computer programmable. As a first step, perfectionism should be avoided as well as overreliance upon a subjectively favoured technique. For example, both linguistic and statistical methods should be considered. The results of automated indexing experiments should no longer be evaluated by a simple comparison with those of intellectual indexing. It seems neither helpful nor realistic to restrict automated indexing experimental work to a very small scope or subject field. Such efforts to achieve minimal improvements cannot be extrapolated nor seem worth the candle.

Independently of the way in which indexing terms are assigned to documents, descriptor retrieval should be regarded as one approach attempting to retrieve documents as well and completely as possible. Free-text retrieval should be regarded as yet another approach; a complementary rather than a competitive one.

Suggestions Concerning Further Development of Free-text Retrieval

Modern free-text retrieval systems sometimes produce more problems than they offer assistance in satisfying informational needs. Under these circumstances, the major requirement would be to reduce the demand for searchers to associate terms for query formulation on his own by procedures offering terms for selection either intellectually or automatically by means of proposals made by the computer. Different methods are described in the literature for the production of such search aids using mainly statistical means (e.g. (4)). Perfectionism should not be required in producing such aids. It is a simple task for the searcher to select appropriate terms and to ignore inappropriate ones. Retrieval systems should be adapted to integrate such search aids and to make their application as simple as possible.

Extensional table for the term MASKING

Statistics: 13 docs. containing MASKING
428 correlated terms

Probability of "co-occurence":

61.5%	PERCEPTION	30.8%	PROCESSING	23.1%	FREQUENCY
61.5%	VISUAL	30.8%	RECOGNITION	23.1%	HEARING
46.2%	BACKWARD	30.8%	STIMULUS	23.1%	PITCH
46.2%	MASK	23.1%	AGE	23.1%	SCHIZOPHRENIA
38.5%	CONDITIONS	23.1%	AUDITORY		
38.5%	SUGGEST	23.1%	COMPARED		
38.5%	TARGET	23.1%	DISCRIMINATION		
38.5%	TARGETS	23.1%	DURATION		
38.5%	THRESHOLDS	23.1%	EMPLOYED		
30.8%	OLDS	23.1%	FRENCH		

EXAMPLE TAKEN FROM (4) FOR PSYCHOLOGICAL ABSTRACTS

Proposal for a Synthesis of Both Methods

Automated indexing may be described as a method of stepwise transformation from a single free-text word or set of them to a term of the documentation language. In contrast, free-text retrieval may be interpreted as a stepwise broadening of a basic retrieval query by the addition of adequate terms using the logical operators in order to match the individual formulations of a concept in the documents.

From these two statements, a simple conclusion may be drawn: it should be possible to apply the same dictionary of terms and their relationships, the prerequisite for automated indexing, towards assisting the searcher in free-text retrieval. At the same time, lists of terms and their relationships developed for automated indexing could be updated and improved by subsequent evaluations of intellectually produced retrieval queries.

The above and often-mentioned dictionary considered as a resource for both methods seems to me to be an overriding argument that automated indexing and free-text retrieval should be regarded as complementary approaches to solve the central problems in information retrieval. Further developments in both areas would benefit by emphasizing their complementary rather than competitive aspects.

References

(1) Salton, G: Computers and Libraries - a Reply. Library Journal 96 (1971), No 18, p. 3277-3288

(2) Zimmerman, H: Ansätze einer realistischen automatischen Indexierung unter Verwendung linguistischer Verfahren. Kuhlen, R (ed.): Datenbasen, Datenbanken, Netzwerke, Vol 1: Aufbau von Datenbasen. München, New York, London, Paris: Saur, 1979, p. 311-338

(3) Lustig, G: Ansätze einer realistischen automatischen Indexierung unter Verwendung statistischer Verfahren. Kuhlen, R (ed.): Datenbasen, Datenbanken, Netzwerke, Vol 1: Aufbau von Datenbasen. München, New York, London, Paris: Saur, 1979, p. 339-368

(4) Wessel, A E: Projekt Informationsbrücke: Entwicklung und Erprobung eines zusätzlichen Programmpaketes zu vorhandenen IR-Systemen. Nachrichten für Dokumentation 32 (1981), No 3, p. 107-114

FACT RETRIEVAL IN THE 1980s

Viktor E. Hampel
Technology Information System
Transportation Systems Research
Lawrence Livermore National Laboratory
P.O. Box 808, Livermore, California 94550, USA

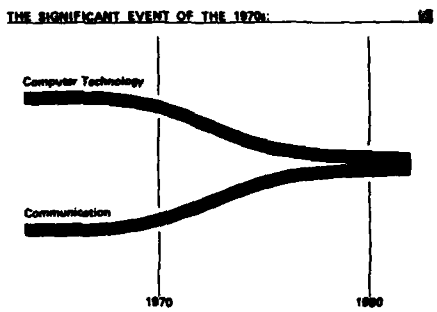
SUMMARY

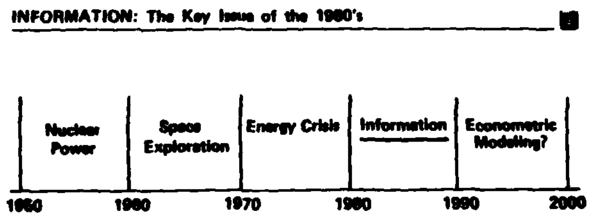
This report reviews prevailing methodologies of fact retrieval in science and technology and makes surprise-free projections for the decade to come: Numeric databases are shown to overtake in size and number the large bibliographic collections. This is expected to lead toward more sophisticated, interactive data analysis techniques with graphical display options. The availability of low-cost intelligent computer terminals, micro- and minicomputers, is shown to make aggregation and post-processing of retrieved information from different sources readily possible. This capability may come into conflict with legal constraints and is bound to affect the traditional marketing of information. It will lead to the extraction of higher forms of intelligence from text and data. The user community is seen to shift from expert information specialists, who act now as middlemen, to the end-users of information. This less experienced user community will challenge the ingenuity of system designers for self-guiding, adaptive, and yet more sophisticated man-machine interfaces. The merging of wide-band digital communication networks with computer technologies will make it possible to interconnect computers, information centers, word processors, and other peripherals, worldwide. Techniques of tabular and graphical fact retrieval are examined. The prospects of fact retrieval by voice, touch screens, and videotext are discussed. The potential of two unusual three-dimensional display techniques, the computer-generated time-resolved integral hologram and the projection of virtual data images into space, are discussed. We conclude by examining the resulting problems and some solutions by example of our experience with the integrated Technology Information System at the Lawrence Livermore National Laboratory.

"Fact retrieval is the identification and use of information about events and measurements by techniques that increase our knowledge, understanding, and ability to simulate and predict social and natural phenomena."

1. INTRODUCTION

The 1980s may well be called the decade of information. Quick access to factual information, the ultimate product of the post-industrial society, has been given a boost in the 1970s by the merging of computer technology with communications.[1] This made it possible for us to generate, validate, and disseminate information faster and cheaper, and has brought it within the reach of any telephone.





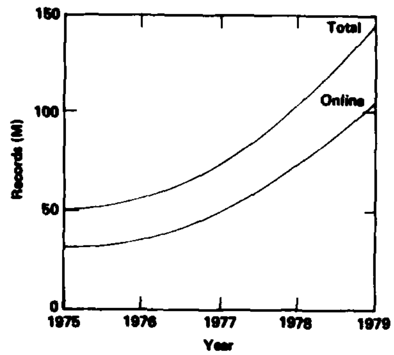
Traditionally, reports and books are used to convey the significance of an event or measurement. The embedded facts are then extracted and compiled in topical lists which serve as concentrated knowledge databases for comparison, evaluation, dissemination, and as a starting point for future work. These compilations of facts are structured to retain the essential attributes of their origin and qualify each individual fact for later use without descriptive text. It should not surprise us to observe that the historic use of computers for storage and retrieval of general information in science and technology has followed the same course.

Textual and bibliographic databases were created first. This marked the beginning of widespread, automated text retrieval services by machine, although digital computers are innately better suited to process numbers. The large bibliographic collections, now fully up-to-date in most fields of interest, represent a comprehensive online index to the books and reports of our civilization. By the end of the 1970s, 1500 discipline-oriented files provided quick reference to 150 million citations of the abstracted literature. Of these, some 450 databases were identified as being available online. They are the "sine qua non" foundation on which we can build the evaluated, numeric data files. Indeed, the relative ease by which text could be edited interactively and printed in different fonts, gave rise to numerous computer-aided typesetting techniques, a great improvement over previous manual and mechanical means to set type, although printing and distribution of the literature continued by traditional means.

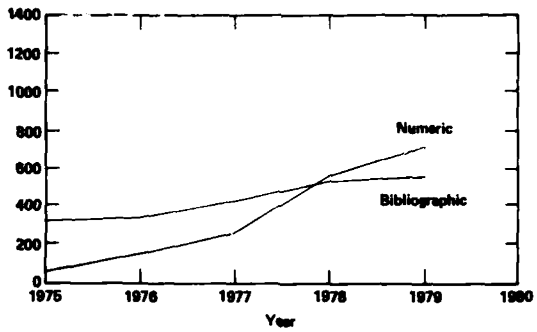
The creation of numeric data files of general interest followed. One of their categories is the huge collection of files created by automated sensors to describe the time- and space-dependent phenomena of demographic and environmental observations. Evaluated data of material properties, which are costly and complex to generate and to retain by computer in traditionally accepted forms, have made their appearance for use by a broader user community only in recent years. The total number of fact files has been estimated to exceed 10,000.

* Work performed under the auspices of the U.S. Department of Energy by the Lawrence Livermore National Laboratory under contract number W-7405-ENG-48.

INFORMATION CONTENT OF DATABASES:



NUMBER OF AVAILABLE DISTINCT DATABASES:



This trend toward numeric data files is expected to receive competition from electronic word processors. The rapid generation of written human communications by machine, and their sharing over wide-band, digital communications networks with other machines and powerful computers, are bound to create a natural mix of factual information, both textual and numeric. To work productively with these facts in large volumes will require the extraction of higher intelligence in concentrated forms. The problem is magnified by the fragmentation of the original information, dissimilar formats, and the necessity to presen' results in concert. This will challenge our ingenuity for innovative techniques of fact retrieval in the 1980s: How can we best establish practical procedures for the creation, storage, identification, validation, and display of the diversity and masses of factual data convincingly and with relative ease?

The immense power derivable from rapid access to accurate and up-to-date economic and technical information has been recognized by corporations and nations. Information management by machine in multinational business enterprises is reported to have increased corporate profits by as much as 30% in 1-2 years.[2] It is our hope that the apprehensions expressed at the recent 7th international CODATA conference in Kyoto, Japan, alluding to the potential prospects of data piracy and information monopolies in science and technology,[3,4] may give way to a mutual and controlled sharing of this emerging powerful resource. This will help developing countries to catch up, increase the productivity of industrially developed nations, and benefit mankind, in general, rather than a privileged few.

In this paper, I would like to review the status of fact retrieval in science and technology as I have had the privilege to observe and practice it at a large national laboratory where advanced computers and communications are commonplace.[5] In the parlance of Herman Kahn, this should make it possible to make surprise-free projections for the years to come.

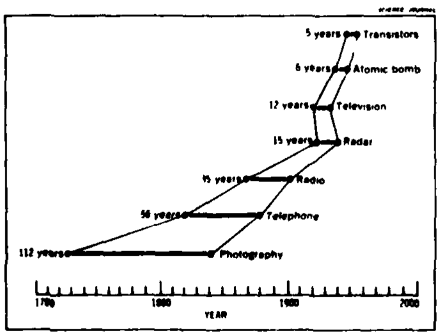
2. THE STATUS OF FACT RETRIEVAL TODAY

There is a saying that we must learn from the past, lest we be damned to repeat its mistakes. Our minds filter bits of information continuously. We compare them with past experience, establish their worth, and remember the essential facts for future use in an unending chain of analysis and synthesis. For these bits of information to be valuable, they must be factual, up-to-date, accurate, accessible, concise, usable, and controlled.

"Fact and Value"

Valuable facts are more than numbers alone. They can be descriptions of events, or well articulated and accredited postulates. For example, the letter written by Albert Einstein to President F. D. Roosevelt[6] expressing apprehension that scientists in Germany may have split the atom, and that it should be possible to build a bomb because more neutrons were thought to be liberated in the fission process than were lost or absorbed, and that the U.S. could preempt them. This letter did not contain a number. But it was so valuable at the time that it started the enormous Manhattan Project, led to the early end of World War II, and ushered in the nuclear age. I am pointing this out to emphasize that fact retrieval need not be limited to numerical data to be of value. Indeed, the extraction of higher intelligence from descriptive text may well be of greater worth. It is capable of projecting our dreams and plans farther into the future than the knowledge of a more accurate event or measurement alone.

The extraction of new insights from large volumes of information, be they text or data, is clearly the challenge of the future. The real payoff in fact retrieval from the literature and numeric data files is, therefore, not only the speed by which this retrieval can be accomplished, but the new insight and understanding to be gained.

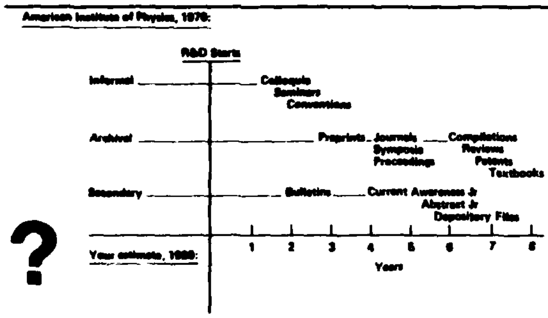


Shrinking time gap between initial discovery and final development.

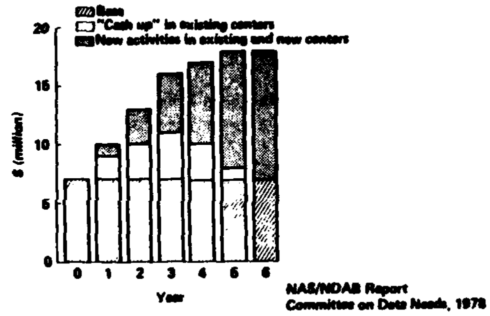
"Datedness and Accuracy of Facts"

Our judgement is only as good as our knowledge! And yet, our information systems today are still hampered by long delays before factual information enters the main stream of computer-aided retrieval and evaluation. Estimates made in a report by the American Physical Society range from 2 years, for announcements of ongoing R&D in Bulletins, to 7 years, for compilations of evaluated numerical data derived from measurements. These estimates were made in 1970 but delays are probably not much shorter today. In the United States, little incentive has been given by government or industry to correct this situation. Other countries, more dependent on the flow of know-how and factual data from abroad than the United States, where 57% of all publication in the energy field originated in 1980, persisted in putting into action deliberate plans for national information systems.[7,8,9,10]

UP-TO-DATE INFORMATION IS DIFFICULT TO COME BY



ACCURATE DATA REQUIRES WELL-FUNDED "INFORMATION ANALYSIS CENTERS"

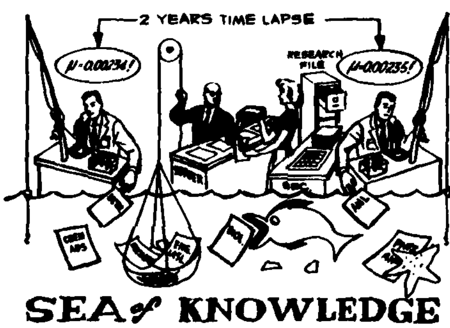


This inadvertent delay before results of ongoing research can be retrieved by computer is further exaggerated in the United States by the inadequate funding of Information Analysis Centers which compile and evaluate the measured data. Recognized data evaluation centers, such as those at the NBS Office of Standard Reference Data, and others operating as members of the Standard Reference Data System, have been working for years with limited and inadequate budgets. The Numerical Data Advisory Board of the National Academy of Sciences, Committee on Data Needs, published in 1978 a report on National Data Needs for Critically Evaluated Physical and Chemical Data (CODAN).[11] It recognized an urgent need for the doubling of support for data evaluation over a 6-year span of time just to catch up with the backlog of existing measured data. The overall national funding for data evaluation in 1977 was only $6.8 million, of which the government provided 90%. Little action of any consequence has been taken to date. Other countries, however, are taking the initiative to capture the transborder information flow and are establishing in new and old areas of R&D their own information analysis centers. We are faced, therefore, with a situation where data are being measured with automated, computer-based equipment in larger quantities than every before but where the users of these data in the United States are obliged to fish them out from the general literature and make their own value judgements of good or bad, to remeasure them, or to buy data from abroad where available.*

"The Cost of Facts"

It is difficult to attach a price tag for facts derived from news or events. Their value is determined by timing and circumstance, as was pointed out earlier with reference to Einstein's historic letter.

Primary publications are priced by page regardless of their value:



Page Range	Domestic Price	Page Range	Domestic Price
001-025	$ 5.00	326-350	$ 18.00
026-050	6.00	351-375	19.00
051-075	7.00	376-400	20.00
076-100	8.00	401-425	21.00
101-125	9.00	426-450	22.00
126-150	10.00	451-475	23.00
151-175	11.00	476-500	24.00
176-200	12.00	501-525	25.00
201-225	13.00	526-550	26.00
226-250	14.00	551-525	27.00
251-275	15.00	526-550	28.00
276-300	16.00	601-up 1	
301-325	17.00		

Add 2.00 for each additional 25 page increment from 601 pages up.

Online bibliographic references that point to the fact that potentially relevant publications took place are sold with even less discrimination at constant unit price, regardless of value or length of their primary publication:

Offline printing	$0.05	—	$5.00 per citation
Online printing or viewing	$0.50	—	$50.00 per citation

Online numeric data are not sold on a unit basis today. As a rule, they are part of a larger file and costs are based on algorithms of Computer Resource Units, storage costs, and connect times, as will be shown later on. As such, they are much cheaper today than the primary publication wherein they are embedded, or the online citations that point to them, even though these facts are the essential and costly part of any document.

* Col. A. Aines, longtime advocate of data management in science and technology by computer has extended a challenge, to any one who could, to compile a report of horror stories where available data were not found in time, or were remeasured at great expense; or even better, where the wrong data were used and caused the failure or delay of costly projects. So far, no one has volunteered. Those involved are too embarrassed to speak up. Perhaps we should ask the retired members of our professional societies to speak to us with courage of their mistakes and lessons learned.

Therein lies somewhat of a paradox: Demographic and time-series data of our environment are massive, and their value may become quickly obsolescent. But, when we look at the cost of accurately measuring material properties in science and technology, we arrive at a very high unit cost. In the previously mentioned CODAN report, a literary publication in the sciences was estimated to cost $45,000 on the average in 1976 dollars and required $4,000 for printing. This was estimated from the time and effort required to do the measurements, assuming a year's worth of research and a median annual salary of about $22,500 with 100 percent overhead. The absolute worth of data in any publication is difficult to judge in the absence of compilations and comparison with other data by expert evaluators. To establish their worth obliges the potential user to familiarize himself with the subject matter, to review the literature, and to arrive at an unbiased value judgement. This points to the urgency and overall cost-effectiveness of authenticated data evaluations.

The cost for compiling and evaluating data is comparatively small. Based on 16 years of operating experience with the Joint Army Navy Air Force (JANAF) Thermochemical Tables, the unit cost of data, representing one material property as a function of temperature and/or pressure, is only $1,000.* Since ten publications are found to support one data sheet in most cases, the investment of $1,000 can capture the essence of $500,000 in R&D expenditures for others to use with confidence.

The scarcity of up-to-date evaluated data compilations of physical and chemical properties probably accounts for their low market value. It leads inadvertantly also to an unaccounted transborder flow of costly primary data. Other countries are making it their business to harvest these factual resources and to market them with advantage!

The Storage of Facts"

Storage requirements of facts by electronic means are assuming staggering proportions. We are committing more written communications, observations, and calculated data to computer-aided storage and retrieval than ever before. Estimates for some of the large data producers and users are:[12]

Lawrence Livermore National Laboratory	50,000	Tapes
Bank of America	400,000	"
Shell Oil Development	600,000	"
Social Security Administration	750,000	"
Exxon Geophysics Logging	1,000,000	"
	2,800,000	"

This represents (at a storage density of 6250 bits per inch for a 0.5-in.-wide magnetic tape of 2400 feet, inclusive of a 30% waste due to record gaps), a conservative capacity per tape of 1×10^9 bits, or an overall requirement for the above sample of organizations of about 10^{15} bits. Only the IBM Photo-Digital storage systems with 10^{12} bits capacity each, came close to filling such a demand; but, all of these mechanical, wet-chemical systems are now retired. Today, as we are entering the 1980s, the projected state-of-the-art for storage technology is estimated to be:

Magnetic	10^9	bits/cm^2
Optical	10^{10}	bits/cm^2
Electronic	10^{11}	bits/cm^2

Of these, only the magnetic tape storage media represented by Automatic Tape Libraries (ATL) are marketed. To appraise the situation at the DOE National Laboratories, and to make demand forecasts for future storage requirements, a survey was made in 1979 with the following results:[13]

Survey Results on Storage Capacity at Nine DOE Sites for 1979.

Site	Tape drives	Tapes	Maximum tape capacity (10^{12} bits)	Disk drives	Disk packs	Maximum disk capacity (10^{10} bits)	Mass archive storage	Mass storage used (10^{12} bits)
Argonne	24	17,120	8.07	128	396	24.0	none	-
Brookhaven	22	[illegible]	1.72	28	56	7.17	none	-
Idaho Falls								
CYBER System	8	18,000	4.68	12	20	2.76	mag tape	8 week
IBM 360/75 System	9	23,600	4.60	26	40	3.2		backup only
Fermi Lab	24	30,000	5.85	20	20	3.2	ATL	
LASL	44	50,000	6.5	70	75	6.0	IBM 1360	2.2
(fixed head)					53	12.7	IBM 3850	2.3
LBL	22	41,500	7.07	24	24	2.82	IBM 1360	3.2
		+ 30,000					IBM 3850	
		archived						
LLL	50	45,000	4.41	110	175	152.0	IBM 1360	3.3
							ATL +	
							CDC 38500	1.0
Sandia/ALB	45	26,000	1.74	77	150	1.51	none	
Oak Ridge	39	36,500	6.54	134	430	39.0	none	
Total	286	325,720	51.2	629	1009	254.4		

* A data sheet may be a single reaction rate constant as a function of temperature. By 1978, 2239 data sheets had been produced in 16 years.

Survey Results on Projected Storage Capacity at Nine DOE Sites for 1984.

Site	Tape drives	Tapes	Maximum tape capacity (10^{12} bits)	Disk packs	Maximum disk capacity (10^{10} bits)	Mass archive storage	Expected capacity (10^{12} bits)
Argonne	24	22,120	11.6	104[a]	15.7	SOA	0.32
Brookhaven	30	8,000	8.08	50	20	SOA	5
Idaho Falls							
CYBER System	14	28,000	8.84	42	72	SOA	--
IBM 360/75 System	13	33,000	10.8	9	4.3	SOA	---
Fermi Lab	40	55,000	55.4	60	9.6	ATL	9.07
LASL	50	50,000	50.4	80	46	tape cartridge	6.5
LBL	14	20,000	11.4	14	2.4	optical disk	2
LLL	17	10,000	10.1	16	35.8	optical disk	100
Sandia/ALB	32	18,000	11.4	120	1.2	ATL	1
Oak Ridge	2	1,000	0.15	4	1.3	tape cartridge	0.4
Total	236	245,120	178.2	499	208.3	---	--

We observe an overall increase of anticipated archival storage by a factor of ten for 1984, and a simultaneous decrease of installed equipment. One of the interpretations offered is an expectation that optical devices should indeed be capable of providing 10 times the storage density now possible on magnetic tape. The results of this survey were used to set specifications for storage devices in the mid 1980s. However, to date we have received no indication from serious respondents. An extract from our specifications follows.

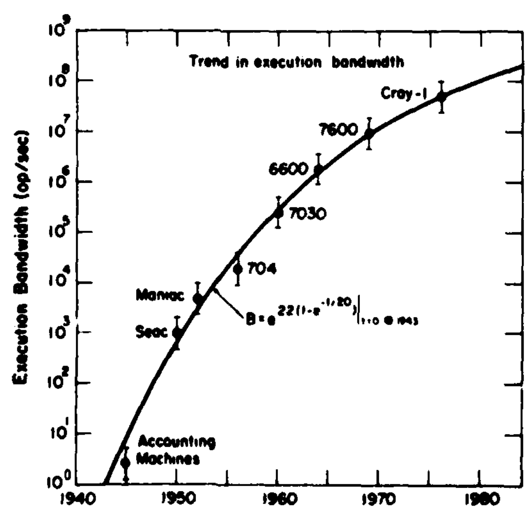
EXPECTATIONS FOR ARCHIVAL MEMORY PERFORMANCE[14]

Property	1980	1982	1984
System capability (bits)	$\geq 2 \times 10^{11}$	$\geq 10^{13}$	$\geq 10^{15}$
Read-write stations (No.)	≥ 2	≥ 4	≥ 10
Read-only stations (No.)	≥ 2	≥ 10	≥ 30
Accesses per hour	~ 2000	~ 8000	~ 30000
Transfer rate (bits/s)	$\sim 10^7$	$\sim 4 \times 10^7$	$\sim 10^8$
System bit error rate	$< 10^{-11}$	$< 10^{-12}$	$< 10^{-14}$
Undetected bit error rate	$< 10^{-12}$	$< 10^{-14}$	$< 10^{-16}$

The rates at which data are being added today are equally staggering: At LLNL, anyone of the four CRAY computers can generate 10^{10} bits of data per hour for a typical 2-D hydrodynamic calculation. As a rule, only 1/10 of that, or 10^9 bits, need probably be saved. A yet larger data source is the Landsat satellites where one pass would produce 10^{16} bits of data, were it not for selective data suppression and compression. But the techniques of file retrieval from these large storage media are not our concern in this paper. They are specialized and are not likely to affect the general user requirements of fact retrieval. Most of our needs in the 1980s will probably be well satisfied by extraction of facts from prestaged archival data files, and from interactions with less powerful, but more numerous minicomputers that have entered our working environment.

"The Analysis of Facts"

Programmatic requirements of the Department of Energy have contributed significantly to the design and development of high-speed computers. Each new generation of the fastest machines has been used for data analysis and to simulate intricate models of nuclear weapons, laser fusion, magnetic fusion, reactor safety, biological/environmental phenomena, and socioeconomic energy predictions. But the rate of performance improvement is slowing down. It took all of the 1970s to gain another factor of ten.[15]



Recent advances in Josephson junctions and the Galts technology suggest yet another quantum jump to 10^9 operations per second in the 1980s. The significance of this expectation is that another generation of even more powerful micro- and minicomputers may also become available for general use. This would accelerate the dramatic changes of the traditional marketing and utilization of bibliographic/numeric data that we are seeing today. The post-processing of retrieved information toward higher forms of intelligence is clearly the challenge of the 1980s.

"Directories to Databases of Facts"

To access and analyze information, one must first know where to find it! The publishing business traditionally prints periodic updates of reference books on topical issues. In recent years, these publications were augmented by printed directories to computer-based resources. The current "Directory of On-Line Databases," by Cuadra Associates lists 770 data files and 135 on-line services.[16] The "Directory to Computer-Readable Databases," updated each year by Martha Williams under auspices of the American Society of Information Science, shows in its 1979 edition 528 distinct data files.[17] The corresponding international directory to computer-based systems with European emphasis, EUSIDIC, published by Learned Information, Inc., lists more that 1280 databases.[18] The somewhat dated report by Westbrook on data sources for materials, cites in 1978 about 300 data files in science and engineering.[19] Unfortunately, these directories are not yet available online even though computers were used to compile them and to print them. This is bound to change in the near future as government and business recognize the value of these master guides to information stores.

"The Tools of Fact Retrieval"

Data storage and retrieval was the goal of the 1970s, and different database manipulation software was built commercially and at universities to bring it about. We distinguish two classes:

1) Data Management Systems (DMS), which permit access to and retrieval from already existing files, usually for single applications. Most bibliographic online information retrieval systems are of this type.

2) Database Management Systems (DBMS), which manage and maintain data in a prescribed structure for the purpose of being processed by multiple applications, independent of storage device class or access method. They organize data elements in some predefined arrangement in a database and retain relationships between different data elements within the database. These systems are commonly used with numeric/structured data under the user's control.

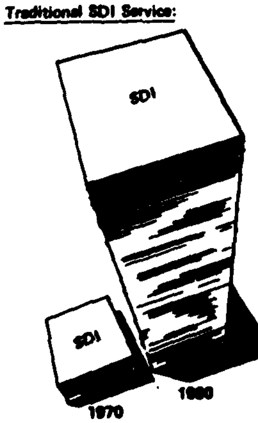
For large volumes of data, like those being communicated to earth from observation satellites, efficient and specialized programs were developed and are usually not suited for generalized applications by casual users. For comparatively small and diversified collections of data, a host of more generalized, less efficient data management systems have evolved. Depending on the logical relationship among data sets and data elements, the major DBMS models now in use are those best capable of working with:

- o simple, COBOL-like 'flat files',
- o hierarchical or tree-like data,
- o graphs and networks, (CODASYL),
- o relational tables, and
- o data well represented by extended set theory models.

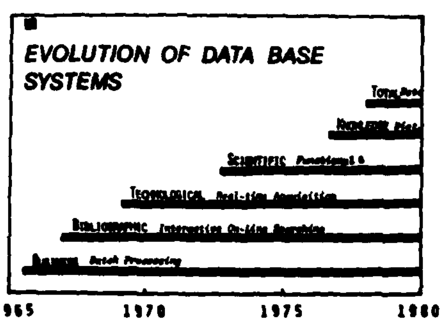
No definite bounds exist among the systems that implement these models. Most systems show considerable flexibility. No concensus has been reached to define the best model for general database work. An overview of their relative merits and historic dependence was made by Fry.[20] But, to those shopping for potential, suitable candidates among the large number of systems, their underlying theory, strengths, and limitations may only be of partial interest. The questions more likely to be asked concern user-oriented features; the types of computers and operating systems on which the systems can be installed; the host languages to which they interface; and cost.

The online bibliographic retrieval systems and services have changed little since their inception in the late 1960s. Only the numbers of online topical data files and citations have quadrupled. Many still seem to believe that the greater the number of citations retrieved in retrospective searching, or by Selective Dissemination of Information (SDI), the better the service. Relevancy, recall, and precision were hotly debated issues and vary with the sophistication of the indexing and retrieval method, and whether keywords, and/or titles, and/or abstracts are scanned. At LLNL we tried and documented several experiments and settled on searching by indexed keywords and words in titles. Abstracts introduced too much noise. Most of the bibliographic systems are patterned after DIALOG (Lockheed) and provide Boolean operators for retrieval of whole words or compound expressions from indexed tables of keywords and authors. Sets can be created and combined to further refine a desired result. Few bibliographic information management systems offer capabilities for interactive algebraic work with numbers. Free-text searching of titles or abstracts is less frequently available. But, regardless of how these systems retrieve their citations, nearly all bibliographic information today is being delivered as a pile of paper. At best, it is printed or flashed on the CRT screen in reverse chronological order of its publication date, requiring the hapless recipient to just look at it!

Traditional SDI Service:



"Look, how much I found for you. . ."
"You can't store it in your mini. . ."
"You can't reproduce it by any means. . .
. . .just look at it. . ."



Retrieval with embedded character strings, weighting of terms, and proximity indicators have become more prevalent with the large legal text retrieval systems, e.g. JURIS, LEXIS, and WESTLAW, among others.

A detailed comparative analysis of numeric/structured data management by means of the commercial systems, or by university systems, is difficult because of the continuous change exhibited by these systems, estimated to number about 65 worldwide.[21] Suffice it to say that most of them tend to be modular with interfaces to procedural languages and to statistical and graphical programs. Excellent seminars on "Comparative Database Management" are being offered each year by the University of California Extension Division, Los Angeles. Comparisons of 21 systems and their capabilities were made by Auerbach[22] and Datapro.[23]

One of the most promising retrieval systems for numeric/structured data developed during the 1970s for general use is the Chemical Information System, (CIS), sponsored by the National Institute of Health (NIH) and by the Environmental Protection Agency (EPA).

CIS is a network of chemical databases equipped with computer programs that permit interactive searching and retrieval from these databases. At the heart of the network is the Structure and Nomenclature Search System (SANSS) which is used to identify a chemical substance, given its name or its structure, and refer the user to all CIS files that contain data on the compound. Any such data can then be retrieved with simple commands from the appropriate node of the CIS network. A different way of using the CIS is to enter experimental data, such as mass spectral peaks, into the system, which will identify the compound at hand, using its files of ten of thousands of mass spectra, nmr spectra, or x-ray powder diffraction patterns. The entire CIS has been developed by cooperating agencies of the U.S. Government and is available in the private sector for use by the public on a fee-for-service basis. It is accessible worldwide through the TELENET telecommunications network. Today, the CIS system is still a collage of files and programs originally provided by developers of each database. However, gradually, a unifying approach is being implemented that will make CIS the most comprehensive collection of physical and chemical data files.

The CODASYL reports,[24,25] published early in the 1970s, identified common and desirable features of database management systems and had a very beneficial impact on their evolution. They provided a basis for the database architecture and Federal database standards.[26] Generalized Database Management systems (GDBMS), tailored to account for attributes and peculiarities of data in science and technology, did not materialize as yet. In most cases they were adaptations of systems used in business. Attempts to point out the special requirements for scientific data were made by the specialists' conference on "Generalized Data Management System and Scientific Information" in 1978, sponsored jointly by the Department of Energy Office of Technical Information and the Agence de l'OECD Pour l'Energie Nucléaire.[27] This was followed by a NASA-sponsored conference on Engineering and Scientific Data Management[28].

For those who may wish to further explore the evolution of fact retrieval during the 1970s, I would like to call your attention to the success of the automated indexing of the bibliographic citations by the Defense Documentation Center,[29,30] the experiments of data "tagging and flagging" still ongoing by the American Institute of Physics,[31] and early attempts at integrating text and numeric data [32]. Additional topical literature is found in the Annual Reviews of Information Science,[33] especially in Volume 14 of the 1979 proceedings of the "Online Information" meetings,[34] and the international conferences on very large databases.[35]

"The Marketing of Facts"

Fact retrieval from commercial vendors costs about twice that from government installations where recovery of expenditures became an operational requirement during the 1970s. In the last few years, the pricing has remained reasonably stable, except for costs of connect-times. We quote from the Cuadra Associates Spring 1981 Edition of the online catalog:[16]

"Pricing policies for access to and use of the online database services are extremely complex. There are a number of components to the prices and they are combined in many different ways. In addition, prices are subject to change with fairly short notice. All of these factors make it difficult to treat the topic in a standard manner. However, there are some general points that can be made. In at least 30 percent of online services, there is an indication that some type of subscription is required for gaining access to the database. These subscriptions range from a few hundred dollars per year to several thousands of dollars. In some cases, the user subscribes to a package, which may include one or more databases and additional services (e.g., consulting). In other cases, the user has several options, each a combination of a subscription price and an associated usage charge.

In general, the major components of the usage prices for online database services differ according to the type of supplier and the type of database (e.g., whether it is a bibliographic database or a numeric database). There are two major groups of policies: one for the timesharing firms and their (largely) numeric databases, and one for the others, covering most of the other types of databases. There are, however, exceptions in each group, e.g., where a timesharing firm offers service on a referral database but prices it more like a bibliographic database.

Pricing by timesharing firms in business primarily for numeric databases, requires for most a monthly minimum (e.g., $100 per month) that is applied if the total usage charges for a given month do not reach the minimum level. The usage charges include the following components:

Connect Time	$1.00 -	$21.00	per hour
Computer Resource Units	$10.00 -	$1.25	per unit (varies with definition)
Disk Storage			varies

These rates can also vary within a specific service, depending on the speed of the terminal being used (e.g., 300 or 1200 baud) and the time of day in which the processing occurs (e.g., prime time vs non-prime time). In addition, the Computer Resource Units charged for a particular database may be greater than the standard timesharing rates charged for other data processing services. This difference occurs either because the database system that is being used is more demanding of resources, or because a surcharge has been added to the standard rates (by a multiplier or additive factor) as a royalty to the producer of the database. In most other cases, pricing by online services is based on an hourly connect-time rate, plus telecommunications costs for network access, if applicable. The hourly

connect-time rates that are cited in the supplier's literature may include the royalty, or the royalty may be cited separately. The range of connect-time rates (including applicable royalties) is from about $25 to $300 per hour. The average is approximately $65 per hour.

For bibliographic and some referral databases, there is an additional charge for offline printing, which is generally based on the number of citations or pages. The range of charges for offline printing is from $0.05 to $5.00 per citation, although for a few databases they may be considerably higher. The average is about $0.15. There may also be charges for online printing (i.e., displaying retrieved information directly at the terminal). These fees range from $0.50 to $50.00 per item. Both online and offline printing charges may vary depending on the amount of information that is printed. In some cases, the use of a service involves a startup fee, which often covers account setup, initial training, and materials. Occasionally this startup fee also includes the cost of a special terminal and/or leased lines to the online service's computer. Many of the online services that focus on the provision of database access also provide volume discounts or have subscription plans that provide for various levels of connect-time rates, depending upon the expected level of usage."

The cost of using the CIS is approximately $45 per connect hour for most components, $75 per connect hour for the major files, notably SANSS and PDSM. In addition to the search costs, there is an annual $300 subscription fee, which is used to defray all storage costs. Costs now levied by other government information centers are similar.

Search times per database require on the average about 15 minutes and are reported to range from just a few minutes to more than an hour. If we consider the 5 million searches, or queries, conducted in 1979 against the 150 million records of computer-readable files in the United States and Canada alone,[36] one arrives at an estimate for the total commercial revenue of marketing information on-line: $125,000,000/year. Actual costs to the buyer who provides the information specialists, terminals, and research facilities, are much greater.

"Other Issues of Fact Retrieval"

There are other issues which have their origin in the 1970s: Wide-band communications, the electronic office, user-friendly interfaces, security, copyrights, and novel ways of fact analysis and display. Because of their significance in the years to come, I treat them jointly by discussing their expected state of technology in the next decade.

3. PROSPECTS FOR THE FUTURE

The rapidly increasing high-speed digital communications are starting to link previously separate communication channels. Voice, data, and video are beginning to serve as an integrated medium for fact retrieval. Their analysis and synthesis on powerful micro- and minicomputers will bring about new forms of communicable intelligence. This, in turn, will increase the value of factual information and introduce controls for its exchange and use.

"The Communication of Facts"

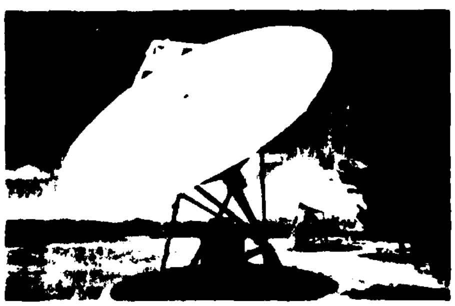
Faster and cheaper computers forced increasing demands for high-speed transmission of digital data in the past decade. Rates increased from 9,600 bps for early telephone lines to 12-14 GHz for satellites approved in January, 1981, by the Federal Communications Commission (FCC) for Satellite Business Systems (SBS). The impact of this change in a relatively short span of time will even be greater when the additional 20 domestic satellites authorized by the FCC last December come into operation. This will permit full integration of voice, data, video, and image transmission. The advantages are bound to affect every-day communication for fact retrieval:

1) Communication costs will cease to be a function of distance.

2) Point-to-multipoint broadcasting will make simultaneous updating of distributed databases practical.

3) Organizational, computer-based networks could be established virtually overnight.

4) The higher frequencies will permit the use of smaller earth-station antennas, 5-7 m in diameter.



At the Lawrence Livermore National Laboratory we now use two antennas with the WESTAR satellites which link the Magnetic Fusion Energy Computer Center (MFECC)[37] network with Princeton University. However, for most of us, it will take several years before the required number of ground stations are up and ready for use, and before the band width for linking to these ground stations can be increased by local hyperchannel networks or fiber-optics communications, among others.

Today, as this paper is going to press, costs per month for cross-country communication between San Francisco and Washington, D.C., are typically those shown below:+

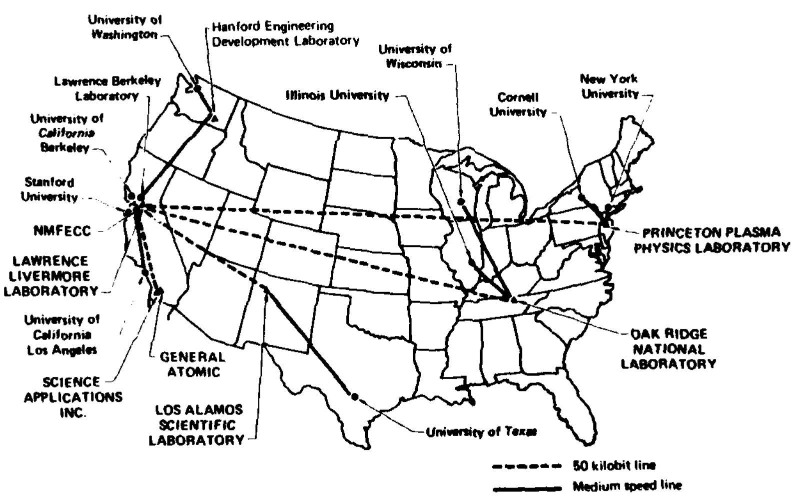
RATE	ANALOG With Modems	ANALOG Without	DIGITAL
1,200 bps	$ 2,140	$ 2,048	Not Offered
2,400 bps	2,218	2,048	$ 2,124
4,800 bps	2,430	2,048	2,338
9,600 bps	2,756	2,048	2,684
19,200 bps	5,942	2,048	Not Offered
56,000 bps	25,490	2,048	11,286

* Vendor sources unknown, conditioning would be required.
** No source available for vendor-supplied Data Service Units (DSUs) required to operate on the D.D.S. network.
+ Private communications with Jack Hibbard, LLNL, Communications Office, July 1981.

The rates quoted are with and without Bell modems. Installation charges for analog communication are $100 without modems, and range from $254 to $639 with modems.

The digital computer networks ARPANET, TYMNET, and TELENET, made their entry in the 1970s. Of these, ARPANET is sponsored by the Department of Defense and was the first large-scale experiment in computer communications by linking various large machines on many university campuses and government installations. Although successful for medium-size file transfer between computers, the 50K-bit lines and packet switching techniques are inefficient for low-speed terminals operating in full duplex. This does not really matter since traffic on the ARPANET is still comparatively light. (Total number of bits moved on the MFECC net exceeded that of the ARPANET already in 1974.[38] Nevertheless, ARPANET became one of the most widely studied and publicized networks.

NATIONAL MFE NETWORK 1981



TYMNET developed quite differently. Its primary purpose was to interface large numbers of low-speed terminals to a relatively small number of time-shared computers operated by Tymshare, Inc. Most of these terminals also required full-duplex, character-by-character interaction with the machine. The inefficiencies echoing characters back and forth were solved by virtual circuits where data from many users shared the same physical record during transmission, so that the overhead of checksums and record headers could be divided among the users. Also, by controlling flow node-to-node, rather than circuit-end-to-circuit-end, there was no need to signal back a message requesting more data. This efficient operation permitted 40 300-baud interactive terminals to be served simultaneously by one 2,400 bps line. In 1972 the National Library of Medicine put the first non-Tymshare computer on the network. Since then, this Value Added Network (VAN), providing automated speed, code, and protocol conversion among terminals and computers, and error recovery by routing records around inoperative paths, has become one of the largest networks with more than 25,000 daily users in 1979, and connections to overseas.

TELENET commercialized the ARPANET concept in 1975 and was approved as an international record carrier. It interconnects to TYMNET and TRANSPAC in Europe, and the Canadian CNCP and TCTS networks. These three networks will continue to have a significant role for ground communication, and interconnect to satellite traffic as required. Existing and planned cable-video connections are expected to assume a significant share of the digital communications market.

The interconnection of these and other national carriers has brought about an international worldwide network. Hilsenrath at the National Bureau of Standards has recently surveyed the field and reports 250 different numerical data systems of which 53 deal with physical and chemical properties.[39] The networks are TYMENT, GTE TELENET, CYBERNET, GE GEISCO, and CYPHERNET.

"Standards for Data Exchange"

The opportunity to communicate computer-based information in large quantities introduced the necessity of standards.

For communications among computers, the X.25 communications protocol is probably one of the most remarkable achievements of the past decade. It has simplified the physical interconnection of host computers over networks. However, the interprocess communications among computers requires either a translation of often incompatible primitive operators, or their standardization. Both approaches are presently in progress.[40,41] The Systems and Network Architecture Division of the National Bureau of Standards has the mandate to study the issues involved and to formulate standards in digital communications. The related problems and opportunities for the then Energy Research and Development Administration (ERDA) were studied and reported by the Working Group on Computer Networking of the Office of Engineering, Mathematics, and Geosciences.[42] Problems and solutions of distributed data mangement and computer networks are reviewed each year in conferences at the Lawrence Berkeley Laboratory.[43]

For communication of facts among dissimilar databases, several standards evolved in the 1970s:

ANSI 239.2 - 1971 American National Standard for Bibliographic Information Interchange on Magnetic Tape.

ISO 2709 - 1973 Documentation - Format for Bibliographic Information Exchange on Magnetic Tape.

ISO 646 - 1973 7-Bit Coded Character Set for Information Processing Interchange.

ISO 962 - 1974 Information Processing - Implementation of the 7-Bit coded Character Set and Its 7-Bit and 8-Bit Extensions on 9-Track 12.7 mm (0.5 in) Magnetic Tape.

ISO 2022 - 1973 Code Extension Techniques for Use with the ISO 7-Bit Coded Character Set.

ISO 2375 - 1974 Data Processing - Procedure for Registration of Escape Sequences.

ISO/DIS 1001.2 - 197x Information Processing - Magnetic Tape Labelling and File Structure for Information Interchange.

UNISIST SC74/WS/20 Reference Manual for Machine-Readable Bibliographic Description.

IAEA/INIS-9 (Rev 1) INIS: Magnetic Tape Specifications and Record Format.

TID-4581-R3 ERDA Energy Information Data Base: Magnetic Tape Description.

These standards deal primarily with bibliographic information. The growing importance of numeric data in the mid 1970s led to the "ERDA Interlaboratory Working Group for Data Exchange," which studied the characteristics of the transmission of numeric data. Based on the previously established standards for bibliographic information, the group proposed a numeric data exchange format, officially referred to as the proposal for an "American National Standard Specification for an Information Interchange Data Description File Format." In 1978, this draft became the substance for the ANSI X3L5 committee, which included several refinements and extensions recommended by international reviewers. The CODATA "Task Group for Computer Use" endorsed the standard and asked UNESCO to have it disseminated in other countries for potential acceptance as an internation ISO standard. Formal transmission of the final recommendation by the X3L5 committee for ANSI confirmation is scheduled for July 1981. It may take an additional 6-9 months before the standard could be officially accepted, printed, and distributed.

"The International System of Units (SI) for Numeric Facts"

Shortly after World War II, in 1954, it appeared as if the United States might have the resolve to change from its English system of units of measurements to the metric rationalized and coherent system of units based on the four MKSA (meter, kilgram, second, ampere) units, plus the degree Kelvin as the unit of temperature, and the candala as the unit of luminosity. Although the United States participated in these international deliberations, extension of MKSA and the adoption of the International System of Units (SI) did not receive final approval before 1976.[44] Public and industrial support came even later. The advantages of only one unit for each physical quantity a well-defined set of unique abbreviations and symbols, and the retention of decimal multiples and submultiples of the base unit for each physical quantity, have made it now possible to exchange numeric databases with less difficulty.

The new SI system of measurement is being adopted throughout the world. Its details are published and controlled by an international treaty organization. However, for a number of reasons, it is inevitable that a few other units outside the system be used with it. It is this additional use of non-SI units that leads to controversy and difference between standards that define modern metric practice. Since a wide variety of metric units have been in use for years in various parts of the world, it is natural that tradition would promote use of these old units in formerly metric countries. For this reason, many European and international standards also recognize a number of non-SI units for use. To protect the new system from degradation, and to cooperate with people all over the world, the SI standard is recommended for all work with numeric factual data.

During the transition period, where technologists are still accustomed to recognize the validity of material properties by their remembered values in former units of measurement, I believe that database management systems for numeric data should offer the option of displaying values in original units of measurement, SI units, or both. It is in this manner that our eyes and minds can be trained to learn, remember, and work with the new SI units with growing confidence.

"Standards for Fact Attributes"

Accredition of facts, especially of numerical data and measurements, requires a "shorthand" notation with attributes which a scientist or technologiest would accept in place of the original descriptive text. The minimum necessary and sufficient set of such attributes has been discussed in the literature. The table below is a representative list.[45]

Attributes of Scientific Data

1. Value
2. Uncertainty
3. Units of Measurement
4. Normalization
5. Validity Domain
6. Method of Measurement
7. Conditions and Constraints
8. Type of Data
9. Source of Data
10. Bibliographic Reference
11. Comments
12. Proprietary Status, Classification, etc.

One of the reasons why numerical databases in the pure sciences may not have come into greater use is precisely the lack of agreement on this topic. This may also explain why a recent survey carried out by the World Federation of Engineering Organizations still lists evaluated numerical property data banks as a desirable, yet less frequently used resource.[46] A number of relatively low-level efforts are in progress to correct this.[47,48] A more authenticated approach is needed. David Lide, Director of NBS/OSRD, emphasizes in a recent publication in Science that standards are urgently needed for critical data and data banks.[49] Use of such standards for the reporting of numeric values for a single datum, and/or a set of data, in the different disciplines, could serve to establish confidence for both data evaluators and users. The NBS Computer Institue for Science and Technology and the Office of Standard Reference Data should document their recommendation in the existing series of Federal Information Processing Standards (FIPS).

"The Copyrighting of Facts"

Computer technology confounded the interpretation of traditional copyright laws. Unlike conventional printing techniques, computer-based systems were capable of storing, processing, retrieving, transferring, displaying, and reproducing works of authorship with ease. In addition, any one of the above processes requires, as a rule, the fixation of more than one copy, (core memory, disk storage, CRT display, etc.) and could thus be construed as a copyright violation. Also, authors and owners of copyrights of works argued that their creations should be patentable, because it is the idea, procedure, process, system, or method of operation that should be protected, rather than the form (i.e., source program, code) in which it is described, explained, or embodied. Section 117 of Public Law 94-553 of October 1976, therefore, did not confront these difficult issues and stipulated that computers did not afford the owner of copyright in a work any greater or lesser rights with regard to the use of the work by computers than those afforded under sections of the common copyright law. The National Commission on New Technological Uses of Copyrighted Works was established a few months later to study the matter. It issued the final report in July 1978, and recommended that Section 117 be removed and/or clarified to provide copyright protection for computer programs and databases.[50,51] This was substantially enacted by Public Law 96-517 in December 1980, with the explicit provision that owners of a software product can copy, or authorize copying or adapting, the work without infringement on the copyright owner's rights if this action either constitutes an essential step in using the program with a machine, or serves archival purposes only.

The recommendations were based on public hearings with particular attention to the problems of copying computer-generated images, e.g., bibliographic citations. The problems of copyrighting numeric data were not explicitly addressed, and some confusion exists as to whether numeric time series data of demographic or environmental observations, and of measurements, can or should be copyrighted. Before I offer an opinion as a user of numeric data, I would like to make the following observation.

The rearrangement of the contents of a copyrighted book, and its printing and marketing in modified form in competition with the original book, would probably be an infringement of fair use, as defined in Section 107 of PL 94-553. A similar situation can be inferred for numeric factual data. Since any information or data with a particular embodiment, e.g., the attributes of uncertainty, precision, normalization, operating conditions, material history, etc., can be protected by copyright, their rearrangement and reproduction in whole or substantial part would probably also be interpreted as an infringement when it affects detrimentally the rewards derived or expected by the owner from the original copyright. This observation, if substantiated, would have significant implications on the three processes by which factual databases in science and technology are created: collection (aggregation), derivation (analysis), and compilation (evaluation). In each case, the authors and owners of these embodiments can request and receive copyright protection for their new works. Royalty payments to the owners of copyrights for the contributing works would have to be resolved in court, if not previously negotiated.

Let us look at some examples. The Department of Commerce makes available magnetic tapes containing numeric data of physical and chemical properties, evaluated under auspices of the NBS Office of Standard Reference Data. These databases are being sold to recover costs, and are copyrighted on behalf of the United States, as mandated in Sections 5 and 6, respectively, of the Standard Reference Data Act, also known as PL 90-396 (1968). Of the approximately dozen magnetic tapes now available from the Department of Commerce, only half contain material properties. The majority of the numeric data, measured and reported by the enormous, tax-supported research and development program in the United States, are, to the best of my knowledge, unprotected. Yet, they form the foundation on which compilers and evaluators in the United States and abroad build their topical databases which then are copyrighted and marketed without royalty payments to the experimentalists, the publishers of the primary literature, or the United States Government. Compilations of numeric data are thus treated similar to compilations of citations in bibliographies. The difficulties in data exchange and transfer arise only after such numeric compilations come into being and are copyrighted in a manner that is not in the interest of the United States.

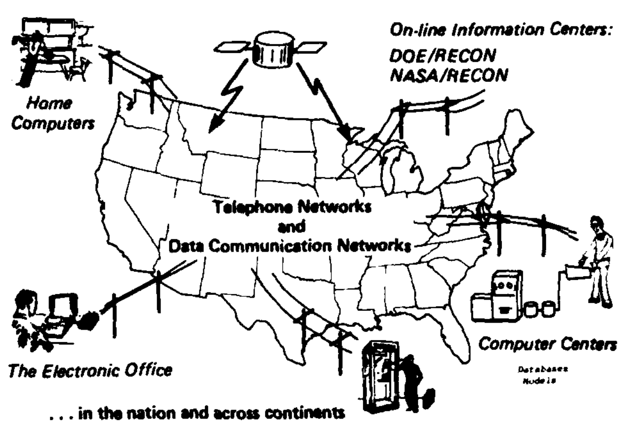
In view of the importance of factual data in the 1980s, the copyright issue of numeric data, and the issue of transborder flow, will undoubtedly be very carefully examined.

"The Electronic Office, Video Transmission, and Electronic Mail"

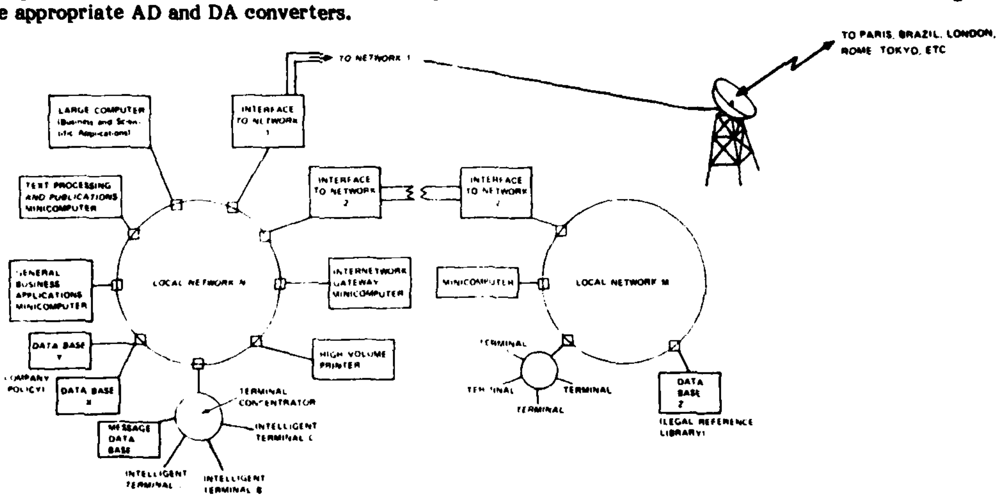
Much has been written on this topic.[52] It is clearly big business when we are given the capability of interconnecting word processors with computers, typesetters, and graphics devices, and can send a resultant illustrated report cross country in seconds. Productivity and creativity are increased. Work satisfaction is enhanced also as secretaries and electronic word specialists now create camera-ready copy that matches the quality previously reserved for publishing houses.

At the present time there are two competing yet complementary telecommunication and computer-based technologies available for transmitting, storing, retrieving and disseminating large volumes of information electronically. One of these, the Videotex (Viewdata) concept, exemplified by such systems as Antiope, Prestel, and Telidon aims at providing general information to a mass consumer and general business audience and is in an early developmental stage. The other, exemplified by bibliographic and numeric information systems, has been discussed before.[53]

ELECTRONIC COMMUNICATIONS ARE LINKING:



Video image and video text transmissions require wide-band communications. These can now be realized locally, but will not be possible cross-country at reasonable cost before sharing of satellite communications becomes commonplace. (One 256K-bit picture requires at least 3.6 minutes for transmission at common 1200 bps speeds.) But, more than 50,000 pictures can be stored on a $4,000 video tape or video disc machine when connected to an intelligent terminal and the appropriate AD and DA converters.



TYPICAL INTERCONNECTION OF TERMINALS LOCAL AND REMOTE PROCESSORS AND LOCAL AND REMOTE SERVICES 52

Electronic mail, usually understood as an expedient means of sending and receiving typed messages and reports, is going to be augmented by voice message systems. Voice input, although now available with limitations, will take longer to implement but is certainly going to become one of the modes of man-machine communication, together with touch-screen command selection as marketed, e.g., by Control Data Corporation in their PLATO system.

The conversion of video text into ASCII text by means of Optical Character Reader techniques is still to come. But, when it takes place -- and it is technically feasible now -- we will experience a total integration of:

Speech ⟷ ASCII Text ⟷ Video Text

The translation of printed text into speech is already being offered by the Kurzweil Reading Machine for the Blind.[54]

"Encoding and Decoding of Facts"

Communications in the 1980s will require protection from eavesdropping and abuse. It is against the law to tap a telephone line. However, the upcoming swarm of communication satellites which is starting to blanket large geographic territories with data and video transmissions, virtually invites anyone with a parabolic roof antenna to "listen" in. User identifications and passwords, if transmitted in clear form, could be compromised. Governments, businesses, and banks have long ago learned to protect their interests by encoded transmissions. The flurry of sales of $2K to $15K roof antennas to harvest the multichannel free broadcasting of television programs will probably be short-lived.

Future users of satellite communications in science and technology will have the advantage of being able to chose from a variety of proven hardware/software combinations to protect their interests. I would like to draw your attention primarily to the public-key cryptosystems that have been proposed toward the end of the 1970s and are likely to have a significant role in the coming years. Their security emphasis has changed from statistical uncertainty to computational complexity. They have developed from conventional private-key cryptosystems to public-key cryptosystems, providing instant privacy and two-way authentication. Some of the essential observations, as summarized by Abraham Lempel, are:[55]

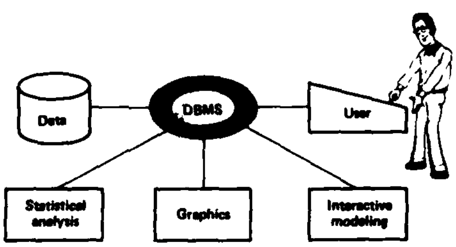
Cryptology and the computational power of electronic computers, once mainly concerned with the making and breaking of secure military and diplomatic communications, have now become a major concern of the public at large. In the age of ever-growing computer data banks and electronic fund transfer, one cannot overestimate the importance of encryption schemes which provide adequate protection against unauthorized, often remote, access to stored data, render the data unintelligible to unauthorized listeners over a publicly accessible communications link, and incorporate a digital signature which can serve as a reliable two-way authentication. These formidable goals were set up to satisfy real market demands, answered in part by the Data Encryption Standard (DES). This is the official NBS scheme to be used by Federal departments and agencies, as well as others, for the cryptographic protection of computer data. Those critical of the DES argued that computers in the early 1980s, rather than toward the end of the decade, would have the power to break the DES code in a day. Alternate schemes have, therefore, become more attractive.

The concept of public-key cryptosystems, introduced in 1976 by Diffie and Hellman, envisioned a system for private communication that employs a public directory in which each subscriber places a procedure E to be used by other subscribers for the encryption of their messages addressed to him, while keeping secret his corresponding decryption procedure D. The existence of such a system would enable instant secure communication between subscribers who have never met or communicated before. For example, if subscriber A wants to send a private message M to subscriber B, he looks up E_B in the directory under B, and transmits $C = E_B(M)$ in the open. Only B can decrypt C by applying his secret D_B to C.

One of the major shortcomings of currently practiced cryptography--the DES, as well as the new public schemes--is the lack of proof that any of these schemes are, indeed, as hard to break as they are claimed to be. Nevertheless, encoding and decoding of factual data will necessarily become a way of life in the 1980s.

"The User Interface"

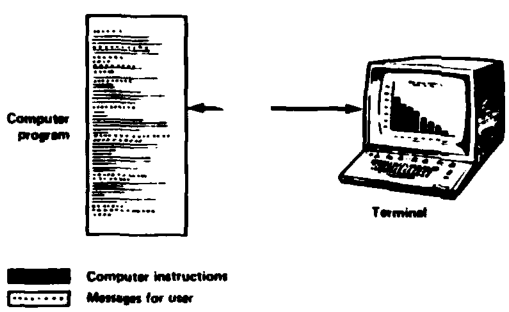
From the preceding review and projections, it becomes apparent that modern technology can link virtually any dissimilar pieces of hardware and software into integrated systems of higher purpose. Differences in format and standards cannot be avoided and do contribute to less efficient operation of the whole. But, they really do not pose a lasting hindrance. Economic pressures cause them to conform. The essential key required to unlock the enormous potential of stored information and factual data lies in the user interface. It is the mediator between man and machine. It makes the formidable aggregate of powerful computers and communications appear to be human, and it imbues us with qualities of exactness and precision more likely to be expected from a machine.



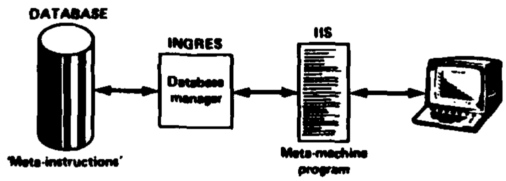
Different methodologies were employed in the past decade to accomplish this goal of translating English-like logical requests into machine instructions, and vice versa. In most cases, the interface has been tailored to serve well one particular Data Management System (DMS), Data Base Management System (DBMS), or some related analysis and graphics package. Although modular in nature, interfaces and their man-machine intercommunications usually became an integral part of the programs they served. Thus they were seldom capable of being extended to other systems, or flexible enough to accept the never-ending demands of an active user community. (DIALOG today is virtually the same as invented more than a decade ago.) References to one or the other approach can be found in publications and proceedings.

Here, I would like to report on the performance of a unique interface, the META-MACHINE,[56] designed and developed by LLNL in collaboration with Control Data Corporation. It serves as an extensible, flexible, and practical interface for the integrated Technology Information System described toward the end of this report. It translates pragmatic English commands into meta instructions for an open-ended number of programs which it controls. It can be readily adapted to languages other than English.

TYPICAL INTERACTIVE SYSTEM

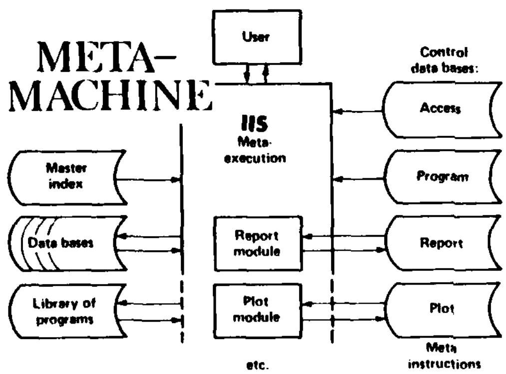


THE 'TIS' APPROACH
Messages and Computer Instructions Reside on Disk in a Database

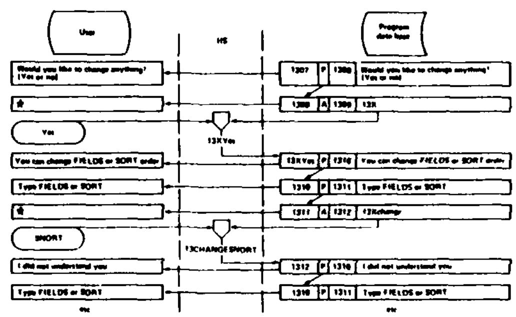


The META-MACHINE is fully self-guiding and supports interactive text and data retrieval, interactive modeling, electronic mail, and networking. It is the central controller for all man-machine and machine-machine communications. It is currently installed on a PDP-11/70 machine at LLNL, using INGRES as the relational database management system, and UNIX as the operating system (2.25 M bytes of core and 1 B bytes of disk storage).[57] However, INGRES and other subservient programs are invisible to the user. Unlike in other interfaces, all man-machine communications are deposited in an INGRES relational database and can therefore be changed in real time, online, without recompiling.

INTEGRATED INFORMATION SYSTEM



META-MACHINE (SYMBOLIC EXAMPLE OF OPERATION)



The META-MACHINE simulates in software a pseudo-computer where sequential instructions are retrieved from a 5-domain program relation table. These domains, consisting of a primary address, a state function code, a forwarding address, an execution string, and a functional clustering attribute, can be arranged and manipulated by the System Administrator with standard INGRES database management commands. Meta instructions in the execution string contain prompting statements to be forwarded to a user, confirmations, command strings for the underlying database management system, or instructions for other programs.

The user is prompted by a uniform command language. His requests are matched against the available options in the hashed primary address domain of the TIS program relation, and are exchanged and assembled into corollary instruction strings for whatever program is to be executed. The transmittal of the prompting strings to the appropriate user at a terminal, or the transfer of an instruction stack to a program, is carried out by a 40,000-byte program residing in core. The 17 state functions determine the destination and type of string delivery. The following figures illustrate this functional procedure. During man-machine interaction with the user, while the command execution strings are being assembled, partial transmission to the target program, local or remote, can be made in anticipation of the full command string. In this manner, and by anticipating permissible attributes and parameters through the functional command cluster attribute, considerable speed can be gained.

The META-MACHINE approach uses one data management technique for access to all information, data, and executive capabilities. This includes control of access rights by user for each of the major databases, the inidividual relations within each database, the user commands, the preformatted reports and graphical displays, and especially the information about other centers to which it automatically connects.

The flexibility of the META-MACHINE has been demonstrated by the successful integration of bibliographic, numeric, project-oriented, administrative, and budgetary information or data files in one system. The report writer, exceeding the ANSI COBOL74 capabilities, is also driven by instructions from the META-MACHINE. Using this unified technique, the report writer was written in 2-3 man-weeks. Delivery of a report writer programmed by conventional techniques had been estimated to require six months and a minimum of $30,000.

The extensibility of the META-MACHINE was demonstrated when in three days we converted an electric car performance prediction model from batch to interactive use. This model, developed by the Transportation Systems Research program at LLNL for the DOE Office of Energy Systems Research, prompts the user for different scenarios: vehicle type, battery selection, time period, driving cycle, etc., and compares the calculated result with the performance of vehicles equipped with internal combustion engines. Fifteen other models are now in use for different vehicles such as hybrids, using flywheels, batteries, and other energy storage methods. The user of the interactive models can enter his own technical values if he wishes to explore conditions that are not part of the prepared options. Validation of these models is greatly simplified by having the input data stored in individual database relations from which they can be extracted, manipulated, compared, and plotted. Results of these calculations can be saved for parameter studies and time-series analyses.

Multilingual interaction with users is possible by simple translation of the user command language, in the program database, into languages other than English. The corollary instruction strings to the underlying database management system, or to other programs, remain the same.

Different command languages can be activated by the extensible TIS database language by translating the instructions from the available INGRES command language into those of a different, local or remote, database management system or execution program. Where additional options are offered by a new system, additional prompts are appended to the program relation with the appropriate addresses and states in the first domains. In this case, the user interaction strings remain the same.

Electronic mail has been augmented by integrating the local and ARPANET directories into one mail system with many options. Automatic file transfer between word processors and TIS has been used since 1980.

Networking became operational with the incorporation of the NBS Network Access Machine software into the TIS system. Fully automated and transparent connections are made to several host machines over the ARPANET and by telephone dial-out. These, include DOE/RECON, NASA/RECON, MACSYMA at MIT, NIC at SRI, LBL-UNIX, SERI, LBL-VAX, and the DOE Alternative Fuels Center in Oklahoma, among 22 other systems. To connect to DOE/RECON, for example, the user simply specifies the target by requesting: connect doerecon. The progress of the connection is displayed and requires between 7 and 45 seconds, depending on whether ARPANET or telephone lines are used. In some cases we activate at the remote host the needed resources immediately, in which case the target name is made identical to the resource, e.g.: connect macsyma.

At the present time, we offer over 64 major information resources that describe in a hierarchical manner material properties for energy storage materials, technology characterization data, systems data that are aggregates of components used in conjunction with energy storage applications, and a number of interactive models. Econometric files were added to permit market penetration studies. The results can be viewed in over 300 predefined tabular and graphical display formats.

In summary, the META-MACHINE and its implementation for the Technology Information System offers administrators and project staff a central focal point for communication and information management, more fully explained in its capabilities later on in this report. Conceptually, it functions as the basis for a stand-alone, intelligent gateway computer with capabilities of linking any user to local or remote sources of information, permitting him to extract and analyze the results, and to share them with others. To accommodate the larger traffic in communications and personal information management, we are planning to replace INGRES by installation of the directly compatible Intelligent Database Machine, IDM-500 by Britton Lee, Inc.[58] This is also expected to off-load the CPU of the PDP-11/70 and to improve throughput by at least one order of magnitude.

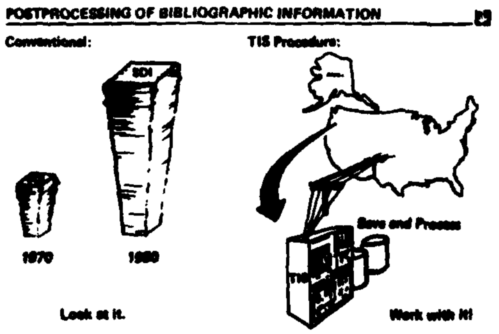
"Extraction of Intelligence"

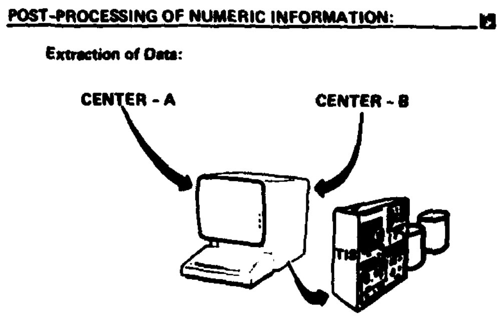
Facts, by themselves, do not convey understanding or knowledge. Their comparison with past experience, correlation with other data and/or predictions, and display in multidimensional color images, however, can be meaningful. As we are approaching the 1980s and are faced with a very large volume of factual data, it becomes increasingly necessary that we filter the data and extract new insight.

In the full-text retrieval service, NTIS has initiated a new type of automated indexing of citations with their Total Microfiche SDI Service, also known as "Selected Research in Microfiche" (SRIM). Since the fiche are being updated quarterly, the subscriber has always a reasonably up-to-date and comprehensive listing of full-length publications in his field of interest, indexed by author, subject, corporate author, contract grant number, accession report, and title--a really remarkable service making good use of the photographic storage medium. This should serve many users well in the 1980s.

In the bibliographic online information services, both Federal and commercial citations are still being delivered to the end-user "en masse" as raw material. It was pointed out already in 1971 that much could be done with the retrieved results while they were still on the host computer. These, and related aspects, were discussed in some detail at the forum on Interactive Bibliographic Systems held at Gaithersburg, MD, October 4-5, 1971, under the chairmanship of Charles Meadow (cf. pp. 173-174):

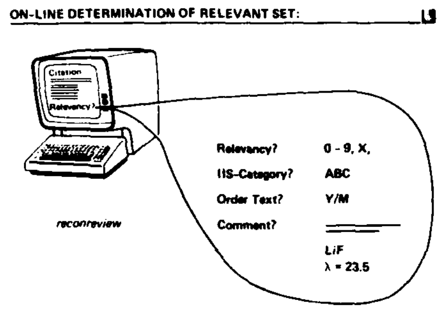
- o Indexes by author, subject, category, etc.
- o Statistics by country, organization, etc.
- o Cross-correlations of data elements
- o Graphical display of results
- o Text analysis

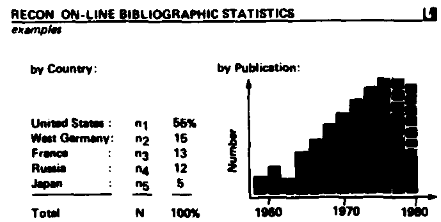




TIS SIMULATES A SUPER-SMART TERMINAL!

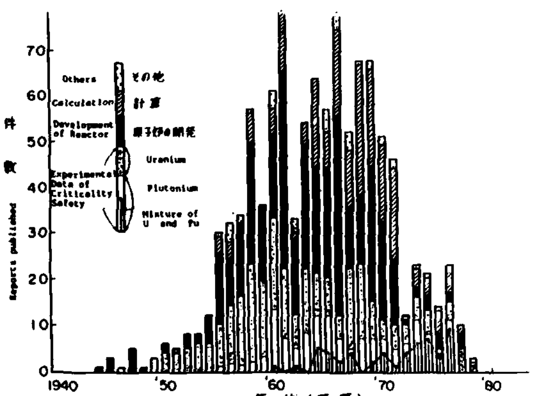
Of these, the first is always used in books. Indexes are sorely needed in bibliographic online information retrieval. At LLNL, we use the capabilities of the integrated Technology Information System, and offer online, interactive commands by which the above options can be carried out by the user immediately after completion of a search. The technique has been adapted especially to DOE/RECON and will be extended to other databases and information systems as required and where permissible. The user can thus create, review online, and annotate his personal or programmatic bibliography and/or library system.

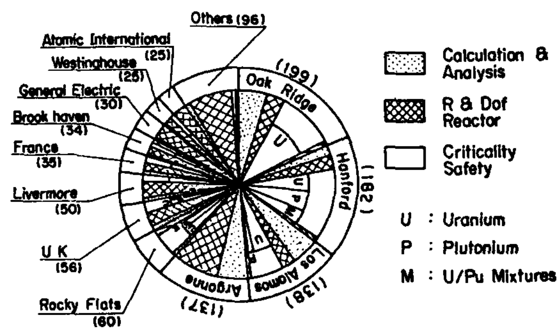




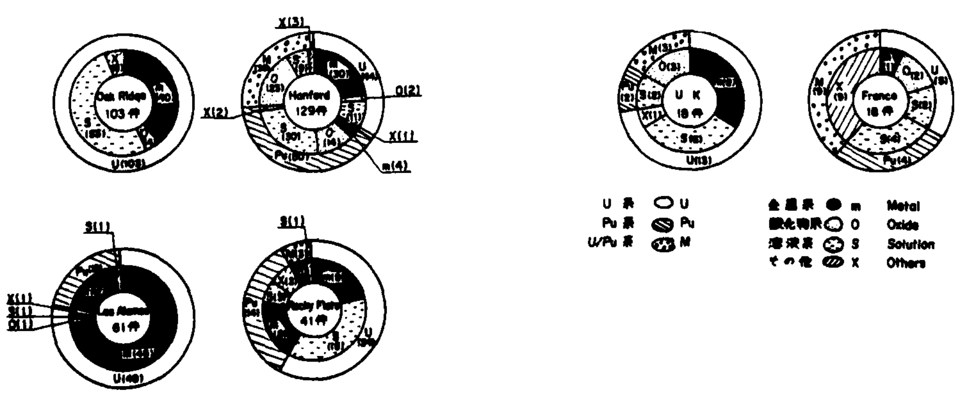
Text analysis, although not new, has some fascination because it permits us to recognize significant migration or cross-fertilization of new ideas in R&D work. Let us take for example the laser field. When one creates an authority list of all single and multiterm expressions derived from titles and abstracts, one arrives at a reasonably stable, slowly changing body of descerptive terms. Newly appearing terms can easily be set aside and used to mark citations for closer inspection. It is thus possible to find the citation where laser beams were first used to weld the retina in an eye. In other words, one can filter out those citations that somehow do not follow the common pattern of previous descriptive indexing or word usage. They contain either typographical errors or, potentially, literary pearls. Other examples quickly come to mind. The techniques for doing this type of analysis have been known since the time when inverted tables, or secondary indexes were introduced for machine-aided look-up of facts. They could be used in the 1980s as a new means of SDI service, signaling unusual and different publications.

I would like to show an additional example of higher intelligence recently extracted by Japanese technologists from a bibliography on nuclear criticality experiments published at LLNL last year.[59] This publication was released in three volumes with accurate computer-generated concordances. It served as the basis for graphical extraction of higher intelligence, e.g.:





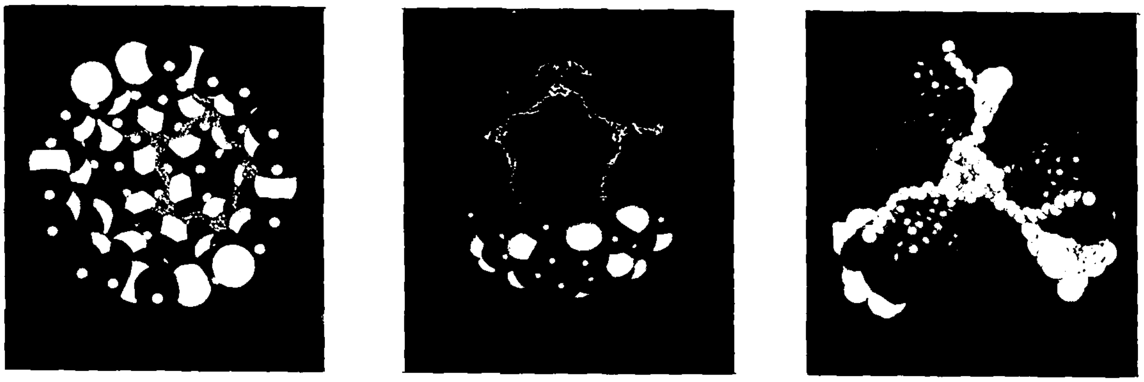
One of the complex pie charts shows the number of reports published from each organization. Another depicts the type of experiments carried out at the different facilities, and finally, cumulative totals of experiments with different fuel enrichments are shown as a function of time. Although these illustrations were probably carried out by manual inspection of the factual data contained in the bibliography, computer programs could be written to do the same routinely.



The extraction of higher intelligence from numeric/structured data files is common practice. Numbers lend themselves directly to statistical analysis and graphical display. General and specialized programs are offered commercially or are under development at universities and Federal organizations. But, what is needed is an adaptation of some of these powerful tools for general use in a self-guided manner. As mentioned earlier, in the field of standard retrieval of bibliographic information, we see a shift from the information specialist as an intermediary to the end-consumer. The analysis and graphical display programs, however, are more complex and will require in the 1980s information specialists well familiar with the manipulation of numerical data and mathematics. The direct creation of electronic visuals (35mm slides, viewgraphs, films, etc.) by computer is the newest exploding technology for the 1980s.

An excellent example of what can be done for technologists with advanced tools is given by the Integrated Programs for Aerospace-Vehicle Design (IPAD). Early stages of the software are in use at Boeing and several other aerospace firms. The system manages the total flow of information and data for engineering and manufacturing. The National Aeronautics and Space Administration (NASA) is making major strides by sponsoring this development.

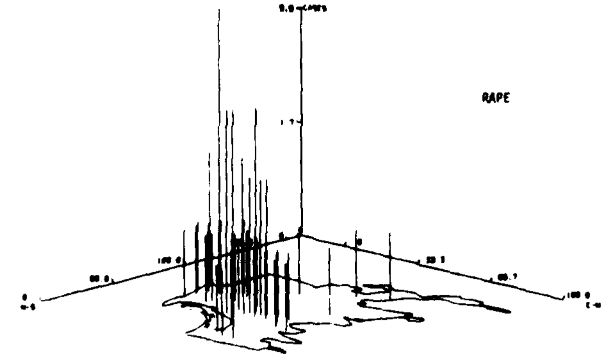
Another example is found in color graphics programs. Beautiful illustrations of double-helical DNA molecules and those of the Tomato Bushy Stunt Virus have graced the front covers of professional journals.[60] They indicate the pictorial power in store for us in years to come. At the present time, these programs are still specialized and not suitable for direct linking with databases. Illustrations are usually produced manually by one-at-a-time adaptations to recently measured parameters, or extractions from an existing database. What is needed here, too, is a user-oriented tool by which anyone with access to the CIS databases, for example, could search, retrieve, display, and zoom in on the molecular structures of interest to him. A number of these sophisticated tools are in the public domain and will hopefully become available in simplified form as well.

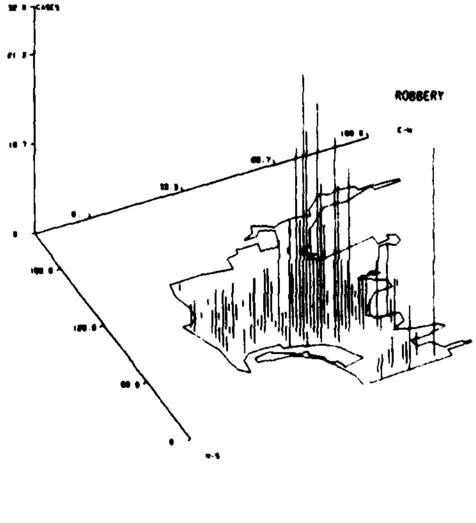


Cutaway view showing inner surface of protein coat of Tomato Bushy Stunt Virus, as represented by computer graphics system under development at National Resource for Computation in Chemistry (NRCC) and LLNL. Each large sphere represents one protein molecule, small spheres are "linking arms" of the protein molecules. Close-up view focuses on three of the protein molecules in the TBS virus, showing arms that link molecules.

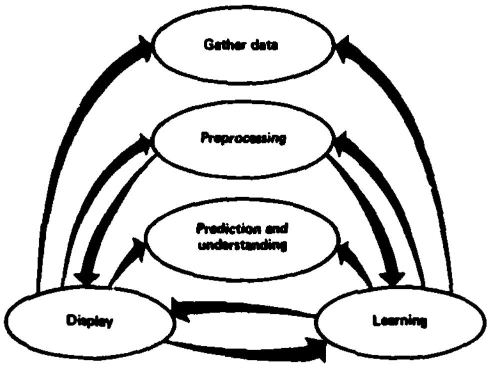
I would like to conclude this section with a remark about pattern recognition. This powerful fact retrieval technique has been used primarily with numeric data, but there is no reason why it could not be used with structured, coded data or even with text. The following examples may serve this purpose:

The San Diego Police Department received a grant in 1975 from the Law Enforcement Assistant Administration (ARJIS) to plan for a computerized data processing system that could serve all of San Diego County. One of the goals was the implementation of an automated crime analysis capability. The Lawrence Livermore National Laboratory was requested to assist ARJIS project personnel under auspices of the National Technology Transfer Program. The results were impressive. Pattern recognition with the LLNL-developed PATTER program could identify correctly those crimes where the success of investigation and conviction was highest. Such an approach can be used for similar multivariate problems that are difficult to solve by statistical or conventional means. In another case, the same pattern recognition program was applied successfully to the primary data describing elements on a wall chart. It predicted correctly the acidic, basic, and graduated amphoteric characteristics of the elements. Pattern recognition has particular importance for work of unusual complexity and offers powerful solutions to problems with limited manpower and budgets.

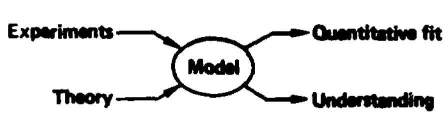




AT LLL WE EMPHASIZE THE ITERATIVE AND INTERACTIVE NATURE OF PATTERN RECOGNITION ANALYSIS –



MODELING



- **Theory is often insufficient**
- **Experimental measurements**
- **Insufficient data for statistical modeling**

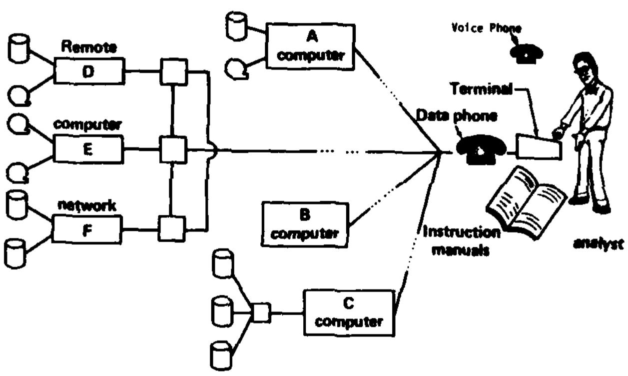
Pattern recognition techniques can be used to construct semi-imperical models

References: C.F. Bender et als.,
Pattern Recognition, Jour. Am. Chem. Soc. 94, 5632, 1972.
Pattern Recognition & Crime Analysis, UCID-17224, 1976.

"The Intelligent Gateway Computer"

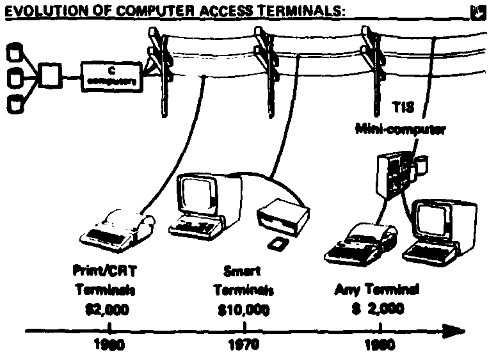
Local office networks are starting to be interconnected with computers, remote information centers, and other resources. The results are not always simple or elegant, but they work and are prompted by the availability of networks and satellites which are removing distance-dependent communication costs. These developments magnify the opportunities for expanding available information resources and computer power enormously. However, to

MANUAL MODE OF OPERATION

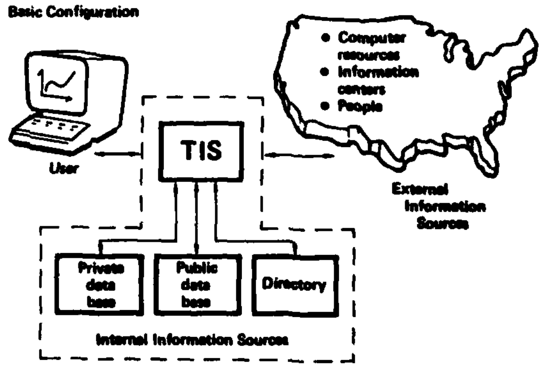


circumvent dissimilarities among different resources and communication protocols, a coordinated approach is needed. Ideally, it should permit a group of users to interact with each other, and with the rest of the world, in a cost-effective and practical manner.

EVOLUTION OF COMPUTER ACCESS TERMINALS:



Basic Configuration



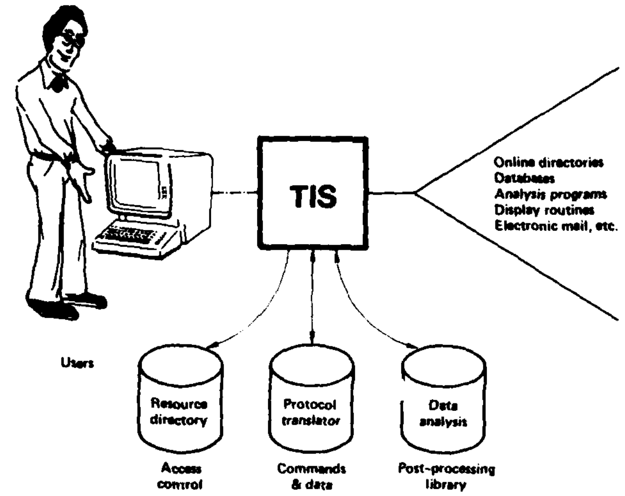
The trend toward this direction has been apparent for some time. Computer terminals developed gradually from mere keyboards to being imbued with microprocessor brains and memories. The communication capabilities of these intelligent terminals, however, are usually intended to connect only to one, or to a few, remote computers or resources. Terminals of this type are offered for $15K to $25K, where some would be considered to be in the class

of minicomputers. In most cases, the manufacturers provide capabilities that are upward compatible only within their line of hardware. This prompts each owner and user of such an intelligent terminal to "upgrade" with even greater investments in time and money without necessarily coming closer to a more flexible and automated solution.

Common communication gateway computers also provide only a partial solution. As a rule, they are not very intelligent. They connect and may exert some control and perhaps do some accounting, but they do not provide translation of incompatible protocols among electronic text processors; they do not permit users to determine their capabilities or allow extraction and saving of information and data from other sources.

What is needed is a stand-alone, intelligent gateway computer. It should contain a master index to the available resources, connect authorized users automatically, translate protocols and formats, permit the aggregation of reasonable amounts of extracted information and data, and offer a resident library of software tools by which these data can be post-processed, analyzed, and displayed.

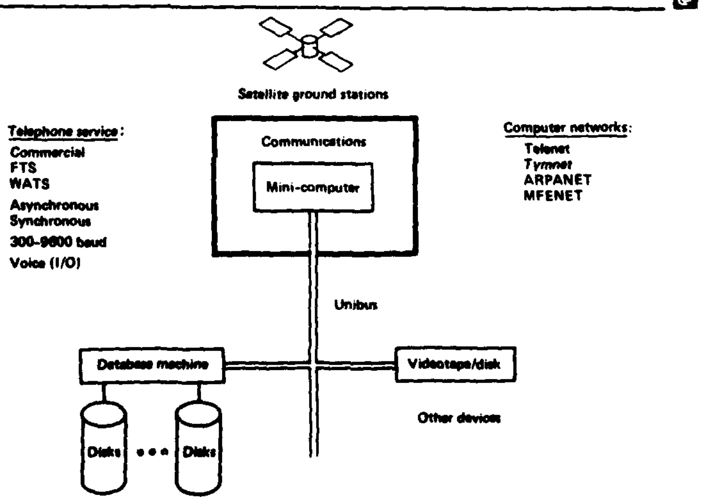
An Intelligent Gateway Computer



At LLNL, we were confronted with a similar problem in 1975 when the U.S. Department of Transportation asked us to study the concept and implementation of a "Transaction Controller" that would permit their analysts to interact with some 26 different computer centers where the statistics of passenger and cargo traffic by ship, air, trains, and trucks are kept. Our work evolved into the design and implementation of the META-MACHINE user interface for the Technology Information System, described previously. Presently, we are generalizing our experience in this field with the concept of a stand-alone, portable system.

The "Intelligent Gateway Computers" would serve a group of users who could retain their old terminals since the intelligence can now be made to reside in a time-shared, interactive minicomputer. A possible configuration might be a PDP-VAX-750 machine, coupled with an IDM-500 back-end database machine, and a versatile communications front for automated dial-out, dial-in, and network access. Our development work at LLNL points toward this direction and uses the META-MACHINE as the extensible and flexible interface:

CONCEPTUAL VIEW OF A FUTURE "TIS" INTELLIGENT GATEWAY COMPUTER



4. FACT RETRIEVAL IN THREE DIMENSIONS AS A FUNCTION OF TIME

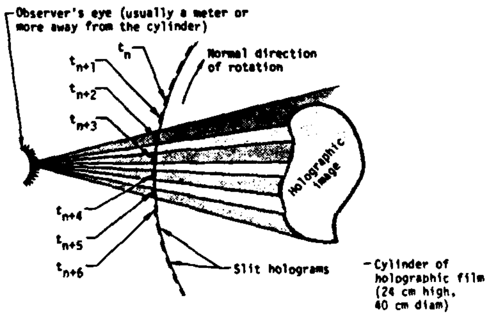
This is probably the ultimate goal of information management. But, as with speech and books, the technology advances first to a state where it can faithfully store and duplicate the information. Management of information, and extraction of higher levels of intelligence, follows later on. Here, I would like to draw your attention to two unusual three-dimensional display methods, those of holography and stero-vision. Both were recently reviewed by Dr. Donald L. Vickers who is leading the graphics group at the Lawrence Livermore National Laboratory. The technical details and illustrations are extracts of his presentation[61] and should give you an immediate insight to the potential of this upcoming technology for your future work.

"Crystallographers, for example, are encumbered with more than their share of the age-old-problem of trying to perceive three-dimensional information. Consequently, they are more painfully aware than most of the irony of living in a three-dimensional world while having to communicate with fewer than three dimensions. Even the computer, with its great speed and ability to solve problems based on data with many dimensions, usually ignores the potential of communicating with the user in at least three dimensions when it comes to graphically presenting the results. Three-dimensional, computer-stored information nearly always is transformed into two dimensions and then plotted on paper, film, or the face of a cathode ray tube (CRT). There are, however, several displays on which three-dimensional computer data actually appear three-dimensional and steroscopic. Two such displays are the integral hologram and the head-mounted display.

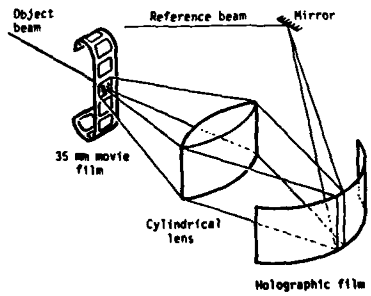
"Integral Holography"

Integral holography is several steps beyond the holography most are familiar with. Many scientists have seen a standard transmission hologram and the three-dimensional image which it forms when illuminated with the monochromatic light of a laser. Many have also seen the cylinder hologram, a cylinder of processed holographic film which, when illuminated with a laser, recreates a three-dimensional image inside the cylinder. As the cylinder is rotated, the object inside also appears to rotate and, thus, one may view both the back and front of the holographic image. Now, try to imagine a cylinder hologram that produces a holographic image which not only rotates as the cylinder is rotated, but which also moves or deforms -- this is an integral hologram. (Refer to figures on the next page.

But even more amazing than a time-varying holographic image is that the source of illumination for viewing is an ordinary incandescent light bulb and not a laser. An integral hologram is not made from a single exposure as it is in an ordinary cylinder nologram, rather, it is made up of 2160 individual slit holograms which are integrated by the observer's eye to form what appears to be a single holographic image. The 2160 slit holograms are made from consecutive frames of a 1080-frame 35-mm black and white movie film by a process described in the following paragraph. The holographic film containing the slit holograms is taped to the outside of a plastic cylinder about 40 cm in diameter such that each slit hologram subtends $1/6^{\circ}$ of arc on the cylinder. As the cylinder is rotated, successive slit holograms come into view, giving to the holographic image an illusion of motion much like that in the 35-mm movie from which the integral hologram was made. The 1080 frames correspond to a 45-second movie shown at a rate of 24 frames per second.



Schematic view from above showing how various slit holograms, though separate in time, form a single "integrated" holographic image. When the cylinder rotates, the eye does not notice any separation between the slit holograms.

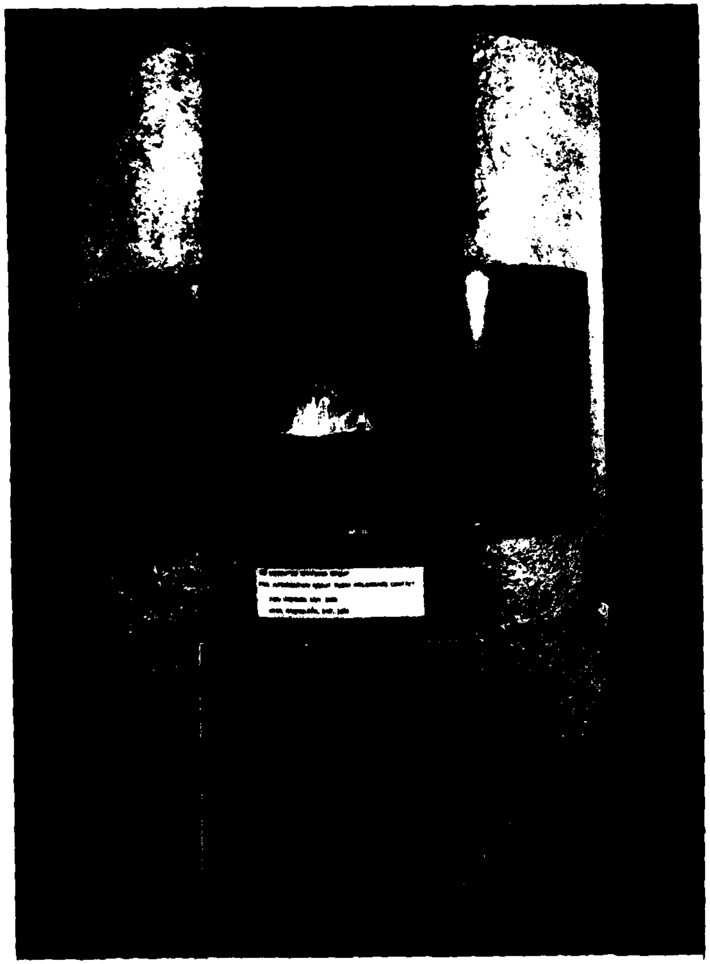


Schematic diagram of the main elements used by the Multiplex Company for making integral holograms.

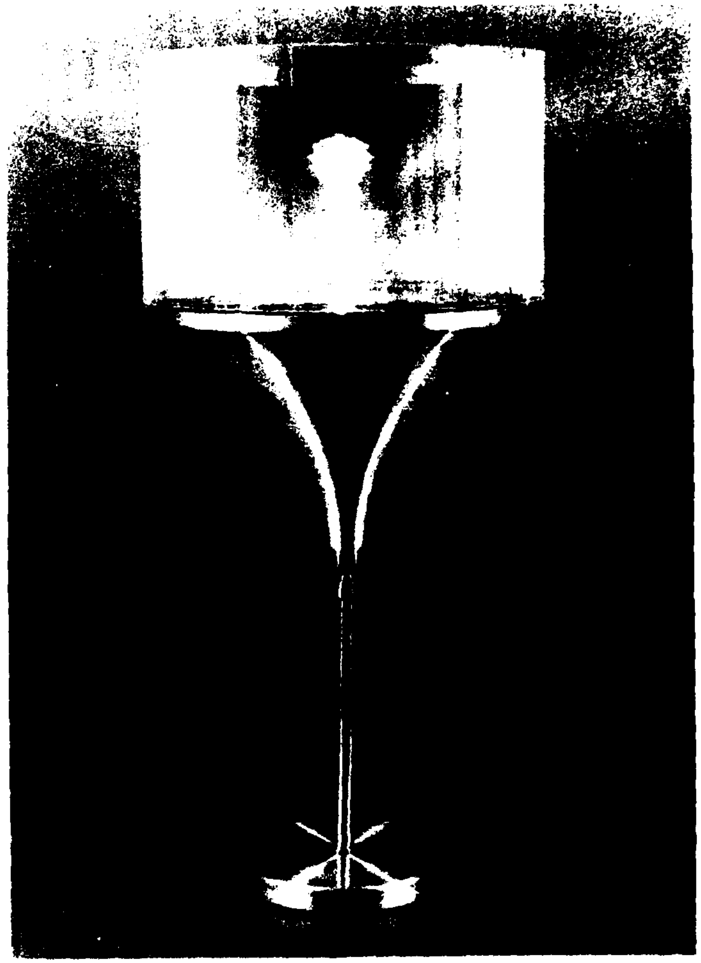
The slit holograms are 24 cm high and 0.7 cm wide on the 1.25-m-long holographic film. They are made by shining a 7-mW HeNe laser through one frame of the 35-mm movie at a time, passing the resultant beam through a large cylindrical lens, and combining it with a reference beam. The interference patterns thereby generated are captured on holographic film which is advanced about 0.5 mm between exposures, thus causing some overlap among adjacent slit holograms. In the "standard" grade integral hologram, the apparent resolution of the holographic image is improved by exposing each frame of the 35-mm movie twice. This is why the 1080 frames of the 35-mm movie result in 2160 slit holograms. A coarser-looking "proof" grade hologram can be made by advancing the holographic film 1 mm and exposing each 35-mm frame once.

Two factors combine to allow the use of white light rather than laser light for illumination of the hologram. First, a cylindrical lens instead of a diffusion screen is used in the exposing process. Second, the subject of the hologram is a two-dimensional piece of 35-mm film and does not require a laser to "unlock" the third dimension. In fact, if monochromatic light were used one would not see the whole image but just a small hoop-shaped band of it and the band would move up and down showing different parts of the holographic image as one raised and lowered one's head. As it is, the holographic film acts as a diffraction grating for the illuminating white light so that the holographic image one sees contains the colors of the rainbow, ranging from red at the top to violet at the bottom. As one moves up and down while looking at the holographic image the rainbow of colors appears to shift. Two restrictions are imposed on the nature of the illuminatint white light: the bulb must be unfrosted, and the filament must be as near as practical to a single vertical line.

To date, the Lawrence Livermore National Laboratory (LLNL) has produced five computer-generated movies of scientific interest from which integral holograms have been made.[62] One shows a nonrotating disk which, as the



Viewing stand showing the Time-Resolved X-ray integral hologram. This two-meter-high stand is on continual display in the lobby of one of the main buildings at Lawrence Livermore National Laboratory. The X-ray hologram was shown at the NATO/AGARD meeting.



Integral hologram showing the calcite crystal structure displayed in a portable viewing stand. The stand, about 40 cm in diameter, uses an ordinary incandescent light bulb for illumination. This stand and hologram were displayed at the NATO/AGARD meeting.

cylinder is rotated, warps according to the phase change in the beam of one of our large lasers. Another shows a rotating sheet which "grows into a mountain," the shape and altitude of which indicate the density of X rays emitted from a laser-bombarded target. Two of the five holograms are of direct interest to crystallographers. The first shows a rotating tetraglycine molecule which, as it rotates, changes from a "ball-and-stick" to a "filled-space" model. The second represents a rotating calcite crystal showing first the unit cell, then adding a clevage rhombohedron, and finally showing the free rotation of the oxygen atoms about the carbon atoms within the carbonate group, as it actually occurs at extremely high temperatures. Refer to figures on preceding page.

The concepts of integral holography are not new. Both the ideas of making composite holograms and of making, from a collection of two-dimensional perspectives (movies), holograms which can be viewed with white light have previously been published. Also in the literature are articles about how to make full-color holograms. What is new is the working combination of these concepts and the commercial availability of the integral hologram such that any interested researcher may take advantage of it.

Applications of integral holography to the areas of advertising and publicity are obvious. The integral hologram is also potentially very useful to both scientific education and research. Imagine the benefit of showing students not just a rotating three-dimensional model of a molecule, but a rotating three-dimensional holographic model that also shows how molecules and atoms combine during a chemical reaction. Already integral holograms have been used with resounding success as visual aids for technical papers in the area of chemistry, crystallography, and computer science. When used as an aid to help the researcher better understand the structure of some molecule or crystal he is working with, the integral hologram is more accurate, more versatile, less expensive, and requires much less effort to create than most complicated ball-and-stick models.

Up to now the integral holograms we have dealt with have been limited to 360° views. (Actually the tetraglycine and calcite integral holograms may be considered to have 720° of view since the molecules rotate twice for every rotation of the holographic cylinder.) Researchers, using a process which requires a monochromomatic point source for illumination,[63] made an experimental integral hologram on a piece of holographic film 70 mm wide and about 7 m long which they spiraled around a plastic cylinder and were thus able to view a much longer movie sequence. When the process of making an integral hologram from a movie becomes more automated, it should be possible to use source and take-up reels which move a full-length movie's worth of holographic film around a rotating plastic viewing cylinder.

Fact retrieval by means of integral holography for most of us may still be some time in the future. However, we recognize that these advanced capabilities are coming up and have to be linked with the data in a user-oriented environment. This is where the producers of data and information specialists come in. Currently, there is probably only one place in the world with the facility for making a white-light integral hologram.*

"Stereo Vision of Facts"

We just learned how computer-generated integral holograms permit us to view complex data as three-dimensional movies. But we remained passive observers of a virtual, tantalizing imagery. To remember these images, to explore the meaning of their valleys and peaks, we still would have to use our minds to ponder their implications from different points of view. Would it not be nice and informative to be able to walk into the data, being able to touch them, to "feel" their irregularities, and to smoothen them by hand. We may have done some of it in an abstract sense mathematically, but must have envied the sculptor who can project the images of his inward eye into reality and mold it by hand.

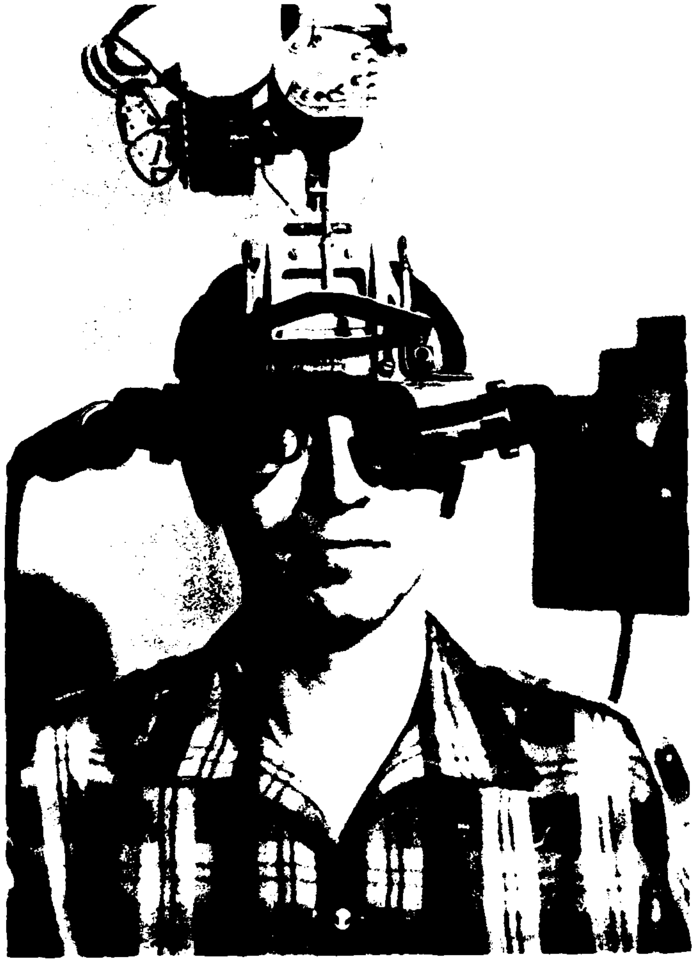
The technology to do this in information management is available. It has been explored during the seventies. I am bringing it here to your attention because technical difficulties and costs that may have prevented its widespread application in earlier years have decreased with the advent of inexpensive, powerful microcomputers that feed on data from larger stores. Stereo-vision of facts is made possible by a head-mounted, three-dimensional viewer. It was first built by Dr. Ivan E. Sutherland at Harvard College.[64]

The display itself uses refreshed CRT technology, but, unlike most CRT displays, this one is worn on the head like a pair of spectacles. Pictures drawn by a computer on the two CRTs are presented to the person wearing the spectacles, the observer, as a virtual image which appears to be made of glowing wire and is superimposed on the observer's field of view in such a way that it seems to float in space. The viewer and a virtual map of the United States are shown on the next page.

The computer-generated, or "synthetic," objects can be programmed to remain stationary as the observer walks around, between, or even through them. They can also be made to change in time as the observer stands still or as he evokes a response by his body motion or controls. Both a wand and light-studded gloves have been used to reach out and interact with the synthetic objects, allowing an observer to touch, connect, deform, erase, and even to create them by "drawing" in space. The unique three-dimensional environment created by the head-mounted display comes from the smooth teamwork of many special pieces of hardware. Refer to Figures.

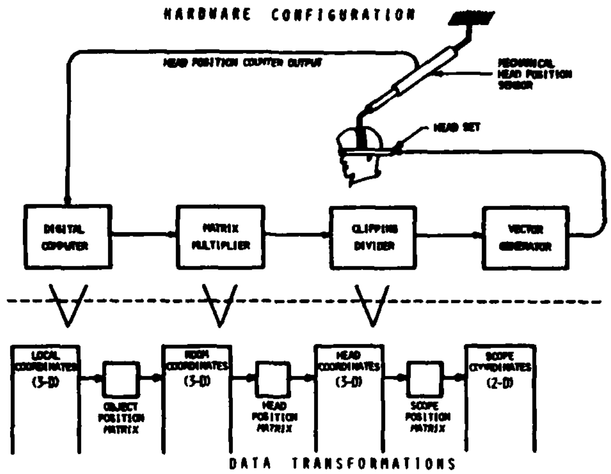
The head set which actually presents the synthetic object(s) to an observer has two 2-cm-diameter CRTs mounted where the temple pieces would be located in normal eyeglasses. A picture drawn on the left CRT, for instance, is reflected off a mirror, through a lens, onto a half-silvered prism, and from there into the observer's left eye. The prism allows the observed to see the real objects in the surrounding room plus the virtual images or synthetic objects drawn by the computer. Stereoscopic viewing is possible by sending different pictures to the right and left eyes. In order for the computer to make a synthetic object appear stationary as the observer moves about, the computer must monitor the position and orientation of the observer's head. This is done with a counterbalanced head-position sensor which consists of a 2-m-long telescoping tube attached through universal joints to both the head set and pivotal reference point on the ceiling. The cyclic chain of events from the reading of the head-position sensors to the drawing of the vectors on the CRTs, takes place at the rate of at least 20 times a second in order to avoid flicker or jerking on the CRT. Future improvements may well use nonmechanical sensors to establish the head position of the viewer. The interaction of different components is shown below.

* The Multiplex Company of San Francisco, which has applied for a patent on their process. You may wish to address your inquiries to them or to Don Vickers, LLNL Graphics Group, L-73, Lawrence Livermore National Laboratory, P.O. Box 808, Livermore, CA 94550, USA.



Observer wearing the head set and harness of the head-mounted display. The lower part of the head position sensor shaft and the encoders for the lower universal joint are visible at the top of the figure.

USA map photographed from different angles through a head set eyepiece. The map consists of 649 line segments and was programmed to appear about 1 m wide and 0.6 m high. Data for the map was supplied by the Evans & Sutherland Computer Corporation. Only one view is shown here.



The wand, the most frequently used device for interaction with the three-dimensional synthetic objects, is equipped with four buttons, a switch, and a potentiometer. These signaling devices allow the observer to "tell" the computer whether he is drawing, erasing, or extracting data. The computer tracks the position of the wand by monitoring the length of line attached to take-up reels on the ceiling. These three spring-loaded lines also serve to counterbalance the weight of the wand and the cord attached to it. (Again, wireless position sensors can readily be envisioned to provide mobility to the user.)

It was quickly evident that four buttons were insufficient to do all the things one might like to do with such a wand, so a wall chart divided into quadrants was designed to multiply the effect of the buttons. Each quadrant of the chart corresponds to a different mode of operation, and each mode redefines the meaning of the four buttons. Of course, only an observer wearing the head set could see the confirmation signaled by the computer to the viewer in the form of a virtual cross. Thus the wand allowed interaction not only with the synthetic objects, but also with any other objects in the room which were "known" to the computer.

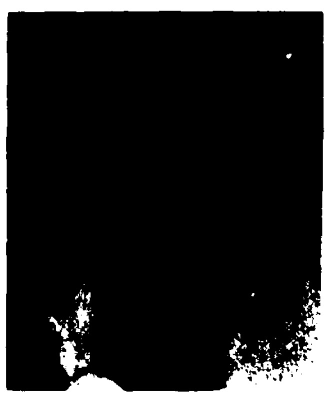
Though the head-mounted display was built at Harvard College, it was taken to the University of Utah by Professor Sutherland just a week or two after its completion. At the University of Utah it served as a research tool for several graduate students. Currently, however, it is inoperable and, though interest persists, funds are lacking. Clearly, the head-mounted display is a prototype. It is not as yet a resource available to present-day cyrstallographers or to scientists of other disciplines.

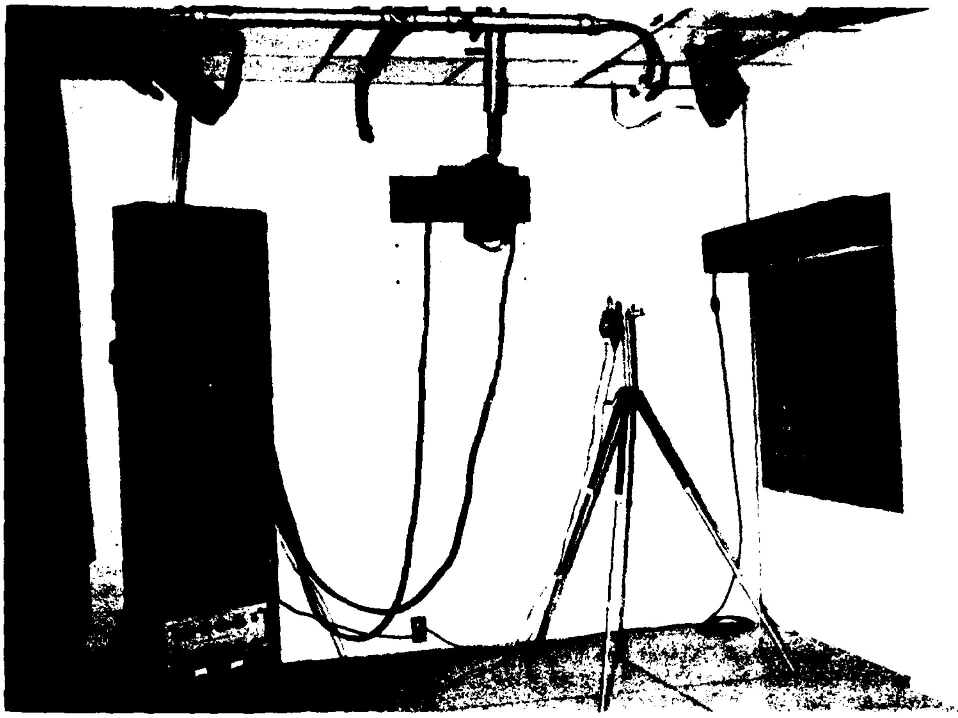
Nonetheless, the head-mounted display has been used with modest success for looking at three-dimensional mathematical functions such as a 3-4-5 Lissajous Figure, for anlyzing three-dimensional electrocardiogram data, and for simulating arterial structure (one was able to walk through simulated arteries). It has been used to "trace" three-dimensional objects, to design three-dimensional, mathematically-defined surfaces and to simulate what a blind person would "see" if an electrode array were implanted in his visual cortex. To our knowledge, it never has been used to look at molecular or crystallographic data.

It is not difficult to imagine a next-generation head-mounted display system which would show colored synthetic objects as solids rather than as outlined drawings. Such a display might be equipped with remote head-position sensing and with telemetered communication both from the wand and to the CRTs, leaving the observer completely free to move around with no wires or harnesses to worry about. Such a display might be connected to a computer which is processing x-ray diffraction data and could be used to look online at a crystal structure model as it emerges from that data. The wand could be used to single out certain atoms in the synthetic image, or to make the image shrink or grow, allowing the observer to walk around it or even inside it.

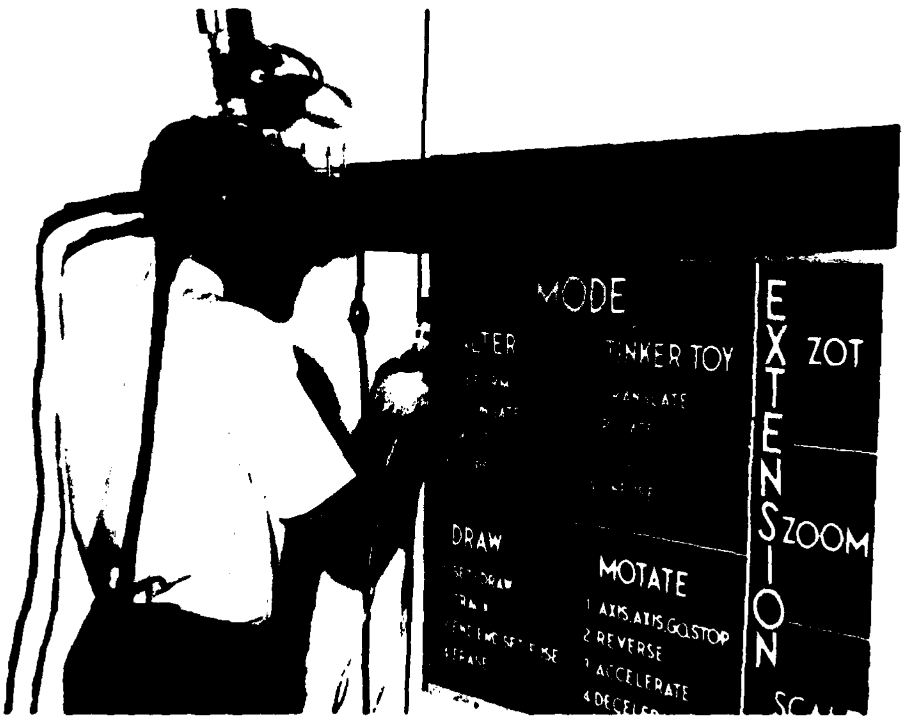
The qualities that make the head-mounted display different from any other three-dimensional graphical I/O device are its abilities: (1) to produce three-dimensional images which are constantly updated under computer control and which change their view as the observer walks naturally about, just as real objects do, and (2) to superimpose a three-dimensional synthetic object on real three-dimensional objects. The data control wand adds the power: (1) to interact with these three-dimensional objects using natural pointing motions rather than by turning knobs or pushing buttons, and (2) to interact with real objects in the surrounding room. And, in spite of being limited to "wire-frame" drawings, the synthetic objects look startlingly real! When this comes about, the finding of facts may be as simple as picking up a synthetic book from a shelf, checking its computer-based index, and reading its virtual pages with all the quickness of the human mind, and projecting their data into space for a personal look and inspection. Computer-aided fact retrieval will then have become an extension of our natural environment."

You can tag, extract, and massage the data with the magic wand.





View of computer head-mounted, three-dimensional viewer, hand-held wand for data manipulation, and control board on wall.



Observer pointing with the wand at the wall chart. The head mounted display and wand allowed user interaction not only with the computer-drawn objects floating in space, but also with real objects such as the chart.

5. THE INTEGRATED "TECHNOLOGY INFORMATION SYSTEM" AT LLNL

Capabilities of the Technology Information System (TIS) provide nationwide bibliographic and numeric database management, interactive modeling, electronic communications, and distributed networking. These capabilities are self-guided and are used successfully by those not intimately familiar with computers. The description of TIS is given here as an example of an operational, intelligent gateway computer, expected to serve technologists throughout the 1980s.[65,66]

TIS is a new-generation, dedicated information machine. Programmatic information is kept on TIS. When additional information or numeric data are needed, TIS connects to other information centers, in an automated and controlled manner. Authorized users simply specify the target name of the desired resource.

In addition, since much of the daily work in R&D is being documented on electronic word processors, we established the capability of linking with several of these machines for transfer of information and data to and from TIS. Translation of formats to WANG, LEXITRON, and QYX word processors is carried out by TIS as required. Now that commercial hardware/software have started to appear on the market, we are planning their procurement to free our resources for other areas.

Analysis, synthesis, and post-processing of information and data are needed to speed up progress, increase productivity, and transfer technology. TIS gives this capability to each user. You, the user, can define and create your own data files, reports, graphics, and communications by activating self-guiding routines. Initially, the results of your work belong to you alone. It requires a permit command to share the data or displays with someone else, a group of co-workers, or to release them for general use.

The system is accessible from any telephone at 300 or 1200 bps, over the ARPA computer network, and soon also over the worldwide TELENET/TYMNET system. FTS and WATS lines are provided for cost-effective use of communications and convenience. Here, I wish to highlight the major capabilities in database management, modeling, and communications. I hope that some of our experience may be useful to you as you are planning your integrated information systems.

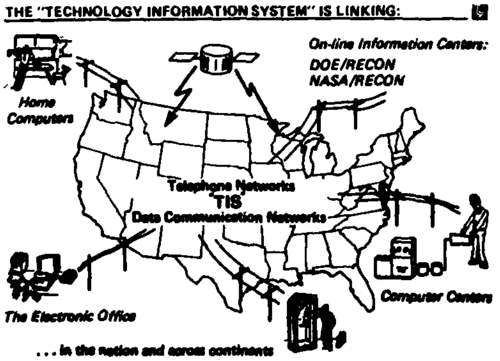
The Technology Information System has been supported by the DOE Office for Energy Systems Research (DOE/ESR). The TIS user community now includes, principally, the staff of the Transportation Systems Research program at LLNL, and that of the Seasonal Thermal Energy Storage program at Battelle, PNL, as well as the DOE/ESR staff. There are about 170 authorized users throughout the country. Electronic communications and the automated access to other information centers is available to all users. This capability is used extensively by the Interagency Information Exchange committee and the DOE Technical Information Center. We are beginning to prototype an Integrated Information Network.

"Database Management Capabilities"

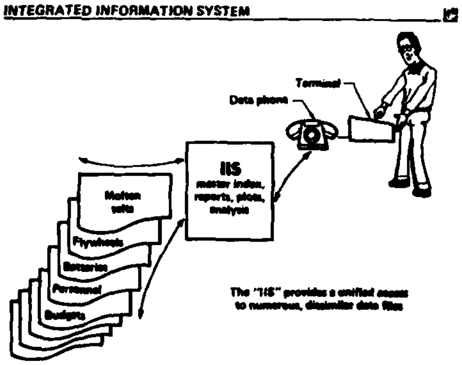
Information is the total of textual and numeric data displayed in a meaningful manner. Most systems specialize in one or the other. Also, most database management systems require a computer programmer or analyst to define the schema of a new database and to load it. Then, when this is done, the database is turned over to the user for retrieval and updating. When special features are required, e.g., more complex reports or graphical output, the services of a programmer are again needed.

The Technology Information System (TIS) offers the traditional database management procedures, but, in addition, TIS has the capability of direct database management by its users without programmer intervention. This permits the use of TIS as an extension of the yellow note pad, or desk calculator. Thus we distinguish on TIS two categories of databases, public and private.

THE "TECHNOLOGY INFORMATION SYSTEM" IS LINKING:

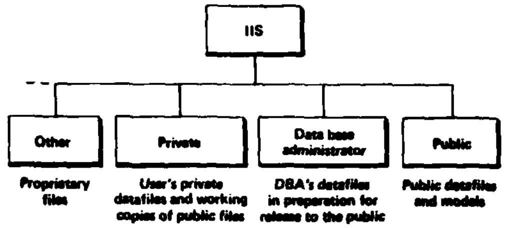


INTEGRATED INFORMATION SYSTEM



The information in these databases is displayed in a hierarchical manner and can be selected with simple specification of an "Option Number," not unlike what you find in some of the popular word processors. However, a database on TIS is a collection of programmatic resources: data files, models, and electronic communications are all different options of the same database and can be tailored to individual programmatic requirements. Six major databases are on the system now. We name especially the database established by the Transportation Systems Research (TSR) program[67] for DOE/ESR, the STES database, and that which contains the installation and technical specifications of thousands of expensive pieces of optics for the SHIVA/NOVA laser fusion program at LLNL.[68] An extract from the display of the TSR database follows:

DOE/ACT TECHNOLOGY INFORMATION SYSTEM



0 AVAILABLE RESOURCES

	1	NEWS
Technical information	2	MATERIAL PROPERTIES
	3	ENERGY STORAGE COMPONENTS
	4	ENERGY STORAGE SYSTEMS
	5	INTERACTIVE VEHICLE MODELS
	6	ECONOMIC ANALYSES OF ENERGY STORAGE TECHNOLOGY
	7	TRANSPORTATION STATISTICS
	8	TRANSPORTATION SYSTEMS RESEARCH
Economic information	9	NATIONAL AND WORLDWIDE ENERGY DATA
	10	NATIONAL ECONOMIC DATA
	11	BIBLIOGRAPHIES
	12	DOE/STOR ADMINISTRATIVE INFORMATION
	13	INTEGRATED COMPUTER RESOURCES
On-line documentation	14	UTILITY COMMANDS
	15	MAIL
	16	HELP

The public databases are intended for general use in support of a particular program. Their information content can be viewed, used, and extracted as required. Temporary changes to these data can be done for display or for ad hoc exploratory calculations by any user. When such changes are made, they are annunciated in the input record documenting a report or model run. Permanent changes can only be initiated by the originators of the data with the help of the Database Administrator.

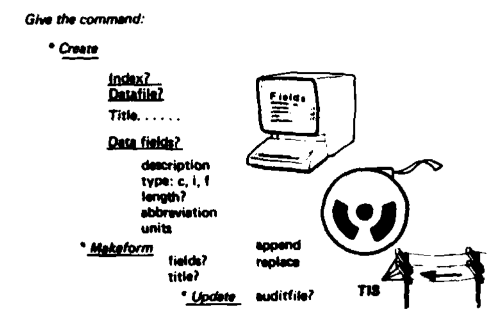
In the private database, we offer the additional capabilities of database creation. The create command, starts a self-guided routine that permits you to establish a hierarchical index for information in your own database system. You can specify and name the data files, and are prompted to describe each data field, indicating whether it will be used for textual data, integer data, or floating-point data; you are then asked to name the units of measurement and select an acronym by which you may wish to refer to the data field in the future as an equally valid name for the corresponding Option Number.

When these self-guided definitions are completed, data can be entered key-to-disk, from menu-driven forms that flash on the cathode ray screen. These display formats can be activated by initiation of the self-guided makeform routines and can be called into action as needed by name. In several cases we have had good results with data input of this type cross-country by secretarial help. The data fields are explicitly called out on the screen. When text characters are inadvertantly entered into a numeric data field, the terminal keyboard locks and signals the error for immediate correction before proceeding. The update command is used to append, replace and replicate data. It provides help instructions for the searching of erroneous records which can then be corrected in a systematic manner. Magnetic tapes are used for input when larger volumes of data are involved. Data can also be transfered over telephone lines or over the ARPA computer network. In the latter case, we can accept data at an effective transmission rate of 36,000 bps.

"TIS" OPERATIONAL RESOURCES:

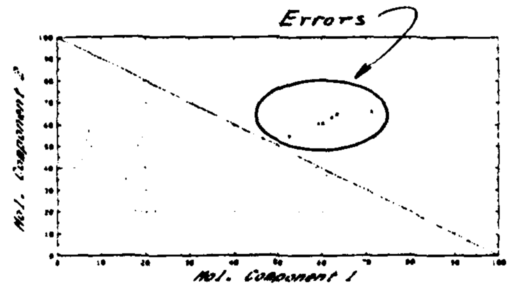
Data bases	64	Technical and economic datafiles
Models:	14	Electric-vehicle models
	1	STES Model
Communications:	64	Communication ports 300-baud telephone lines (Comm & FTS) 1200-baud telephone lines (Comm & FTS) 300/1200-baud WATS lines ARPANET Telephone dial-out Dedicated lines, on site at LLL VOTRAX Voice synthesizer TYMNET/TELNET
Other computers:	22	Automated links to other computers/centers: DOE/RECON, NASA/RECON, DOE alternative fuels (OK), SERI, DOT/TSC, LBL-SEEDIS, MIT-MACSYMA, Worldnews (NYT, UP), etc.

"TIS" GIVES YOU THE CAPABILITY TO CREATE YOUR OWN INFORMATION SYSTEM:



The display and extraction of information or numeric data can be carried out in two ways. First, each public data file comes equipped with one or several preferred display formats, also referred to as reports. These can be activated by name and provide the option to specify Boolean logic for the selection of those records that satisfy numeric and/or textual criteria. Reports can be graphs or tables. You can choose those that fit your terminal. Second, you can create your own reports by initiation of the print or plot commands. These routines guide you to indicate the data file for which the report is intended, the datafields to be printed, summed, labeled, ordered, footnoted, etc. Plots can be seen in black and white on Hewlett-Packard 2648 terminals, or in color on HP 7221 X-Y color plotters. The latter can be used to make viewgraphs directly. These display patterns can be combined with text for reports which are activated by name as an automated sequence of commands. A device-independent interface is being prepared for graphics display on terminals from other manufacturers. In the example below, which shows a scatterplot of the 6200 Eutectic salts, created on TIS from the magnetic print tape to the corresponding NBS publication,[69] we discovered 7 salts where the percentages of constituent materials exceeded 100%. These errors were brought to the attention of NBS/OSRD. This clearly illustrates the power of database management for Information Analysis Centers.

When needed, numeric data can be extracted for later use. They are prepared through the print command and then saved in separate files which serve as direct input to models or as inclusions in electronic mail.




```
Type an Option Number, a Command, or "stop".
*report eutectic where (mol1 + mol2) > 100.

              Molten Salt Eutectic Report Where (mol1 + mol2) > 100.

                                   (1)

                         MOLTEN SALT EUTECTIC DATA

C #  SALT                       MOL %            TEMP (C)    REFERENCES
.........................................................................

209  AlCl3-NaCl-TaCl5           63.1-37.2-1.92   78.00        331
366  AlCl3-POCl3                62-63            119.38      1159
450  AlCl3-KCl                  71-65.5                       688
2001 CaCl2-NaC                  52.5-55                        42    59
                                                               82   183
                                                              295   259
2089 CaCl2-LiCl                 83-64            900.00      1174
4585 NaF-ThF4                   63-64            700.00       148
4694 KF-NaF                     59.5-60                         8    55
                                                               61    62
                                                               74   179
.........................................................................

Type an Option Number, a Command, or "stop".
*
```

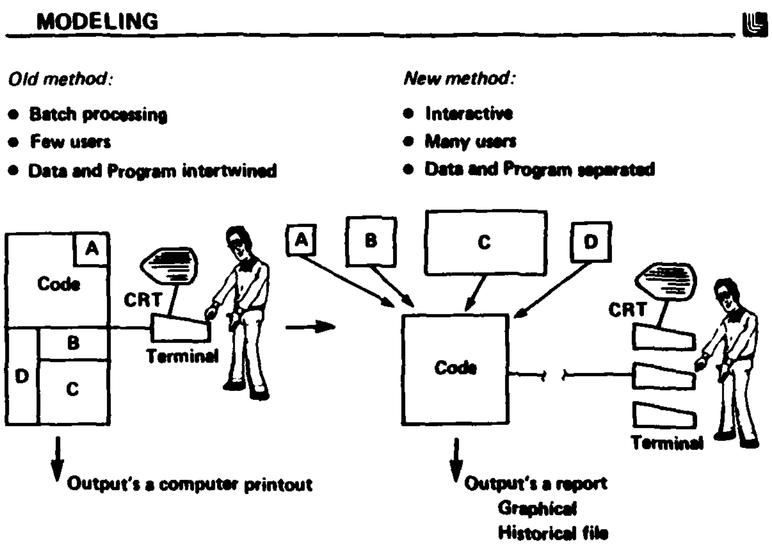
Every report, graph, and sequence of presentations can be used initially only by the creator. A positive permit command is required to share the information with someone else, a group of co-workers, or the user community as a whole. Both in the public and private databases, the user is shown the availability of only those data files, report formats, and display patterns to which he has access.

For those who like to venture into more complex work with data files and reporting, many powerful UNIX utility routines are available. All routines are documented online.

Help is available online for most programs. Commands with many parameter options give help during initiation by the user prior to execution. You may type "help" at each step to receive guidance for the next question to be answered. We offer also online tutorials. This is decribed later when the link command is discussed under TIS communications.

"Modeling"

The execution of simulation models for performance prediction of energy storage systems, or for technical and economic analysis, can be carried out interactively or in the batch mode, in three ways as shown below. We prefer the first mode, which permits separation of data files from the models with inherent advantages.



- o The model may reside on TIS, which controls its input and output.
- o The model may reside in TIS, and be activated by TIS, but prompting is carried out under model control. This is usually the case for imported models.
- o The model may reside on another computer elsewhere in the country, but is controlled by TIS with regard to input or output.

The first method is well represented by the group of models developed at LLNL for performance prediction of electric and hybrid vehicles, powered on various energy storage devices. Originally, these models were used in the batch mode and the code was intertwined with input data. Their use was difficult, limited by computer printouts, and best left to those familiar with their intricacies. To make them interactive, and to prepare them for use by others, the modelers prepared succinct statements that describe the purpose, assumptions, methodology, and limitations of each

model. These descriptions were then used for their online documentation in TIS. Next, the modelers wrote the interactive script that describes each parameter, assumptions, modeling techniques, and limitations. During execution of the model, the user is prompted to select categories and input parameters in a self-guided manner.[70,71]

```
Type an Option Number, a Command, or "stop".
*5

5         Interactive Vehicle Models

   5.1        Electric Vehicle Data
   5.2        Standard Driving Cycles
   5.3        Electric Vehicle with Optimized Battery
   5.4        Battery-flywheel Vehicle
   5.5        EXXON Energy Storage Evaluation Model
   5.6        JPLEV (Elect. Veh. Model)          RESTF
   5.7        Aluminum/Air Battery Vehicle Model
   5.8        CARA Electric Vehicle Model

Type an Option Number, a Command, or "stop".
*
```

```
Type an Option Number, a Command, or "stop".
*5.3

5.3       Electric Vehicle with Optimized Battery

   5.3.1      Model Purpose and Assumptions
   5.3.2      The Modeling Technique
   5.3.3      Basic Vehicle Types
   5.3.4      Basic Performance Levels
   5.3.5      Basic Time Periods
   5.3.6      Basic Confidence Levels
   5.3.7      Battery Descriptions
   5.3.8      Driving Cycles
   5.3.9      Running the Model

Type an Option Number, a Command, or "stop".
*
```

The descriptions and the interactive script for each model were then placed in a small data file and were made part of the overall Transportation Systems Research database. They provide, in this manner, dynamic, up-to-date documentation. The potential user can familiarize himself with any aspect of the model background by selective reference to the appropriate Option Numbers, which describe the origin, purpose, techniques, limitations, and input data files. When he is ready to run the model, he is prompted to select the required parameter categories, or to give his own values. When all answers have been received by TIS, the data are extracted from the individual data files and presented for viewing and confirmation. Ad hoc changes can be made at this time. They affect the run, but not the content of the public database. Following execution in real time, the results are presented in tabular form or as graphical output. We have also devised efficient interactive input methods which the user may wish to use for repeated execution of models.

```
                      INPUT TO MODEL
_________________________________________________________

       THIS COMPUTER RUN WILL START WITH A BASELINE
              ICE VEHICLE DESCRIBED AS FOLLOWS:

 2 passenger car (vt=2)  limited performance      (pl=3)

The projected use is for the  federal urbancycle (dc=8) for the
1985-1990 time period (tp=2) on a  50% confid. prediction (cl=2)

 The above baseline vehicle implies the following data:
 (An asterisk preceding a line indicates that data was
 changed by the user.)

           VEHICLE DATA ( 2 passenger car     )  (vt=2)
 Frontal area of vehicle               (af)   =  18.000 sqft
 Coefficient of drag                   (cd)   =   0.360
 Curb weight of vehicle                (cw)   = 427.000 kg
 Vehicle length                        (vl)   =   9.100 ft
 Power system volume                  (psv)   =   4.300 cuft
 power system weight                  (psw)   = 109.000 kg
 Power-to-weight ratio                (pwr)   =   0.026
 Desired range                         (rd)   = 120.660 km
 Efficiency of mtr/controller       (etamc)   =   0.850
 Weight propagation factor            (wpf)   =   1.300
 Motor weight to power ratio         (wmpr)   =   4.300 kg/kw
 Motor volume to power ratio         (vmpr)   =   0.060 cuft/kw
 Weight of payload                     (wp)   = 136.000 kg
```

```
                      OUTPUT  FROM  MODEL
_________________________________________________________

with the above data, the following design parameters have been calculated

   battery design parameters:
            ragone curve intercept         115.361   w/kg
            specific energy                 88.758   w.h/kg

   vehicle design parameters:
            curb wt.                        959.34   kg
            payload                         136.00   kg
            total test wt.                 1095.34   kg
            vehicle length                   10.11   ft
            battery wt.                     364.14   kg
            motor + controller wt.          122.46   kg
            transmission wt.                 31.89   kg
            road load energy per cycle    1061.600   w.h/cycle

this page of output is held in file named event
Would you like to rerun this model with more changes?
(Type in "yes", or "no".)
```

An example for the second class of models is the EXXON econometric model for electric cars. It is available on TIS, but its prompting is that originally devised by EXXON. Here TIS simply acts as the controller for the model run and provides a convenient means of initiation and execution. Any model which can be compiled and processed on the PDP-11/70 machine can be integrated into a TIS database in this manner.

The third type of modeling capability on TIS is equally powerful. Here the model is executed on a foreign host computer under TIS control. TIS connects an authorized user to the distant computer automatically and activates the named model. An example is the Electric Vehicle Model (ELEVEC) at Jet Propulsion Laboratory.[72] Another example is the "CCC" Thermal Aquifer Model[73] under development at Lawrence Berkeley Laboratory (LBL) for the Seasonal Thermal Energy Storage Program. "CCC" was moved from LBL to the Solar Energy Research Institute (SERI) computer to become part of their solar energy model library. It requires a CDC-7600 and considerable time to execute. In this case, TIS is used to prepare the input file for "CCC" execution at SERI and the retention and analysis of results.

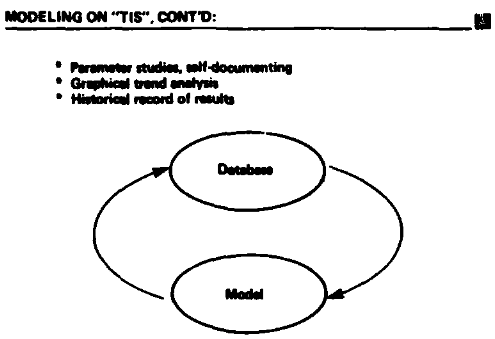
This technique provides a very powerful capability for the execution of models at any site under TIS control. It also offers the possibility of preparing input from common data files, executing the models at different sites under TIS control, and comparison of calculated results on TIS to establish the relative accuracy of models. (Candidates for this effective procedure are the national energy models, which can not readily be moved from their home base.) With this approach, the user can also avail himself of any improved version of the model when it becomes available. Significantly, the models do not have to be translated for use elsewhere or divorced from the creative work of their originators. We are looking forward to an opportunity of using TIS in this capacity.

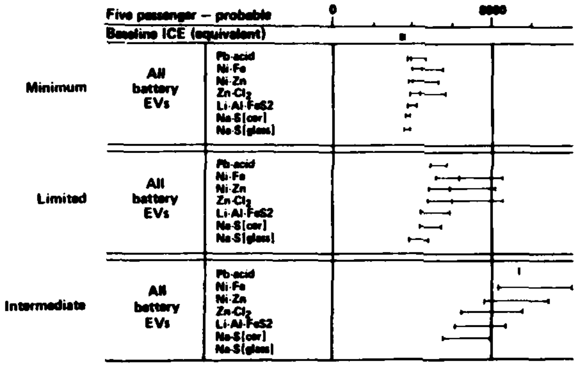
The output from large models can be voluminous and difficult to handle. As a remedy, we envision the extraction of significant calculated parameters from the resulting large output file for viewing and decision making on TIS. The bulk of the output file could be left at the remote host machine. TIS would maintain the index and the bookkeeping. Alternatively, as being planned for the "CCC" model, the output file could be transferred to the LBL computers under TIS control for post-processing on LBL machines and printing on their peripherals. The effective transfer of such large volumes of data requires high-speed communications with transmission rates of at least 9,600 bps.

Modeling requires programming. The major languages available on TIS today are:

FORTRAN IV	APL
PASCAL	"C"
MACRO II	RATFOR
BASIC	DC
SNOBOL	MB
LISP	AS

Several powerful text editors are supported by the UNIX operating system and provide online editing capabilities for a variety of different terminals. Large programs, requiring extensive computer time, can be scheduled by TIS for compilation or execution at night. The results of calculations can be saved in a user-specified library of data files.





Projected initial cost (1977$) of energy-storage automobiles with reduced performance designed to maximize the use of electricity.

Several statistical and graphical analysis routines are available on TIS. We have established, especially in graphics, a number of powerful programs, some of which permit online input in a prompting manner. This pertains to the creation of bar charts, pie charts, and milestone charts for administrative purposes. The graphs can be prepared in color as hard copy or directly as viewgraphs. Once created and named, the resulting format file can be released for use by others elsewhere and printed near-instantaneously cross-country on compatible equipment.

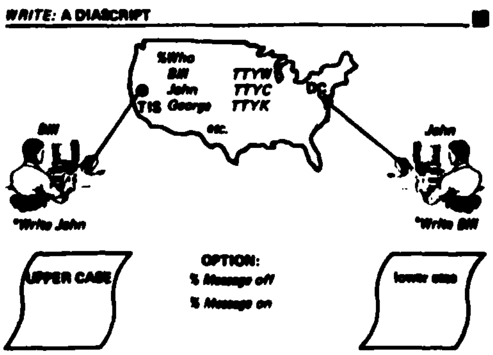
"Communications"

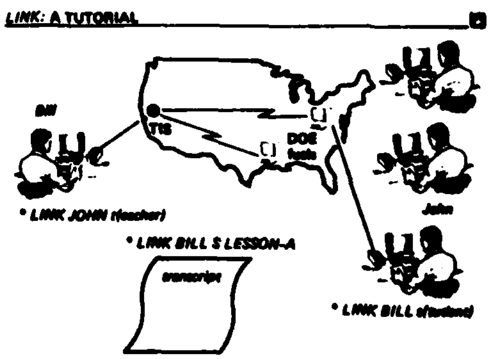
Effective communications are essential for information transfer among co-workers in different time zones. TIS offers the following capabilities:

comment	- a self-prompting routine to send public messages to the TIS database administrator.
write	- a diascript between two users.
link	- provides tutorials for one or a group of users.
electronic mail	- serves the entire user community, inclusive of voting, and the joint preparation of reports.
interconnection with word processors	- permits the transmission of letters and reports via TIS.

Work on an integrated system like TIS invites comments and requests for improvement to the TIS administrator. The self-guided comment routine offers this capability. These suggestions can be viewed by all users of the system and offer an opportunity to see the requests made by others, and the corrective response by TIS management.

The write command is a diascript between two users logged in on TIS at the same time. The write command, followed by the name of the user you wish to reach, prints an alert message on the remote terminal. A similar confirmation from the other side is required to establish communication. Typing then takes the place of a dialogue. A signal can be typed to indicate the end of a question or statement, inviting the response, and so forth. (If you should be doing serious work on an editor or in graphics, and do not wish to be interrupted, you can issue the message off command to silence any disturbance.)



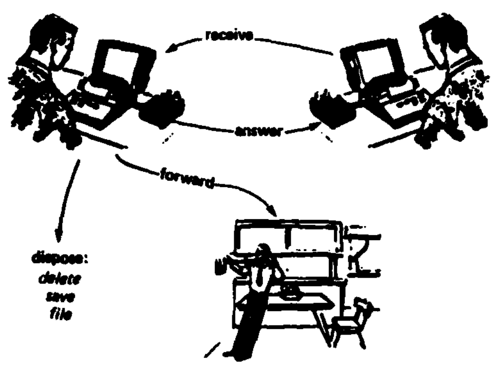


The link command is used for tutorial purposes. By previous agreement, it permits any two users to work together. One user becomes the teacher and works in the student's account. A dropfile can be created for subsequent perusal of the joint transactions. This capability is being used by TIS staff for cross-country tutorials. They are especially effective when used with a voice phone, permitting the student to see and hear instructions simultaneously. Arrangements can be made for class tutorials.

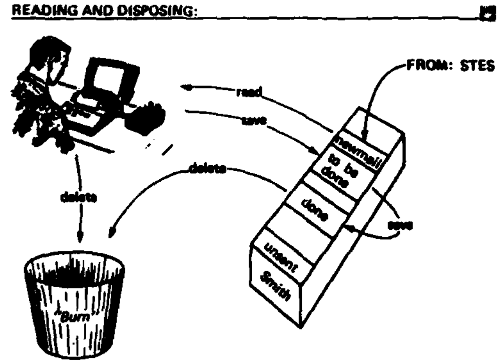
Electronic mail (em) permits you to send and receive messages, to answ... ...d forward mail, to issue group mailings, and to file correspondence in a mail filing system of your own.[74]

Some 26 different options are available to compose and edit messages and reports, correct spelling by reference to the online Webster's dictionary, send blind copies, and check whether an addressee may have already read your mail. Of course, you can delete all mail.

MAJOR OPTIONS OF em:



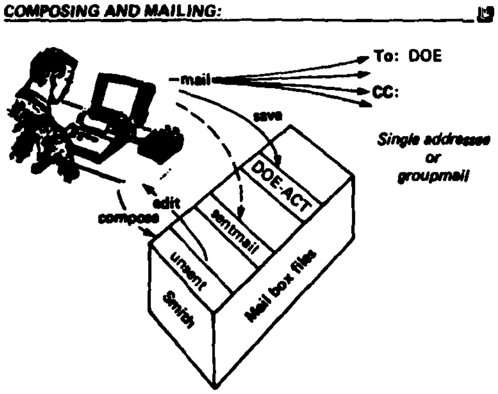
READING AND DISPOSING:



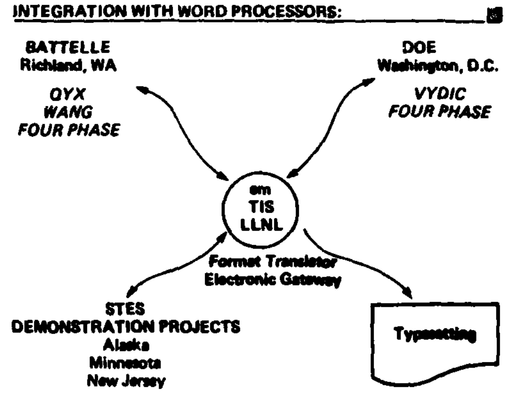
Online help is available for all options. Commands can be executed by their imperative or starting letter. This creates a very user-friendly work environment for beginners and experts.

Interconnection with Text Processors. We established the capability to connect TIS with several word processors: WANG, LEXITRON, and QYX. Connections to the FOUR PHASE system and VYDEC are planned. When used in conjunction with electronic mail, any letter or report typed on a word processor can be made part of an electronic mail message and sent on its way to the destination ahead of any written confirmations. Incompatible control characters among some of the different word processor systems are translated by TIS as required.

COMPOSING AND MAILING:



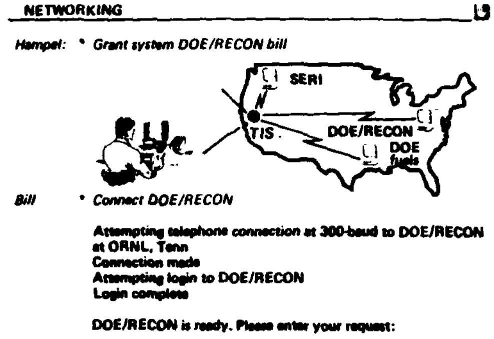
INTEGRATION WITH WORD PROCESSORS:

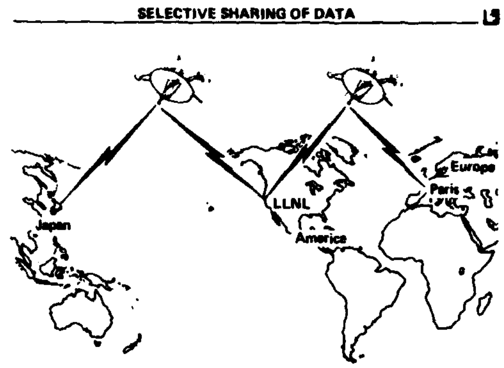


"Distributed Networking"

This is a very powerful TIS capability. It permits connection and use of other information centers and computers in an automated and controlled manner. At the present time, we have provisions for access to 22 other centers, thus vastly multiplying the information content and capabilities of TIS. The Network Access Machine software (NAM) used on TIS for this gateway function stems from earlier work by the NBS Computer Institute for Science and Technology.[75,76]

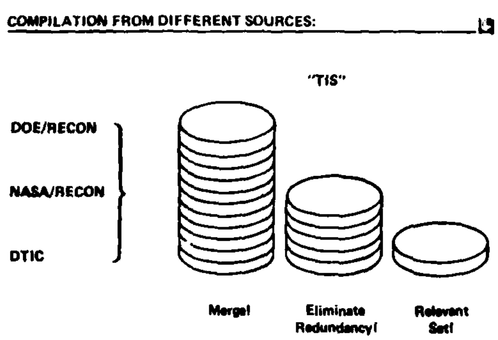
To establish a connection, the arrangements require only one contract with TIS. Individual users on TIS are then granted access as needed for the duration of their work. Audit files keep accurate records of all transactions for accounting purposes.

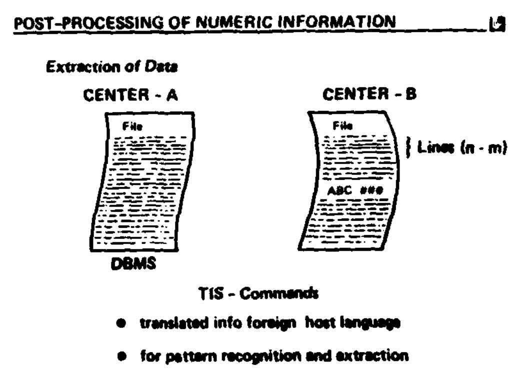




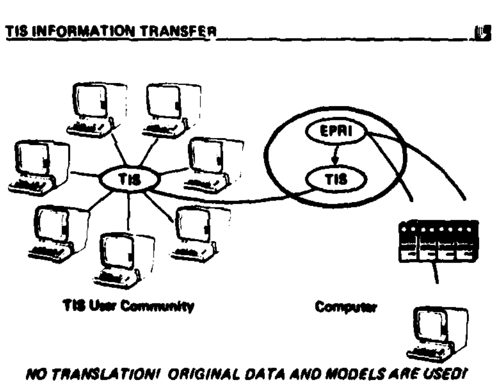
The significant aspect of this capability is that individual users of TIS need not learn the access protocols, passwords, or peculiarities of the foreign host computers. They simply select the information center by Option Number or by target name. TIS does the rest. The power of this gateway approach was recently demonstrated during the 7th international CODATA conference in Kyoto, Japan. With a simple command, "connect DARC," TIS established a computer communication from Kyoto, Japan, via TIS at LLNL, to the DARC system at the Institute of Topology at the University of Paris, France.[77] For extended periods of time each day, interactive graphics of the sophisticated DARC system were demonstrated in Japan, in full duplex at 1200 baud, without any noticeable delay in transmission.

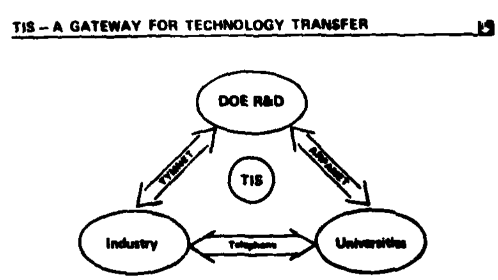
TIS is also capable of retaining the viewed or extracted information, derived from a foreign host, in the user's account for subsequent processing and use where legally permissable. A cogent example is our interconnection to the extensive DOE/RECON information system. All citations retrieved can be placed into a file, aggregated, and processed interactively online for the immediate creation of subject and author indexes, or for topical concordances and text analysis in general. Any bibliographic field element can be cross-correlated with any other. Where required, citations can be complemented with key-to-disk annotations about their relevancy and ranking. Requests for full-text copies can be issued automatically. Citations can be augmented with numeric or descriptive data derived from the reports. Publication statistics can be shown graphically, online, and enhance further the insight possible from the retrieved information. This opens new vistas for extraction of higher intelligence from descriptive text in science and technology.





We expect to have similar links soon with NASA/RECON and with the unclassified Defense Technical Information Center (DTIC). These files are in the public domain and could be used to establish comprehensive, well-indexed bibliographies, now carried out more laboriously. This capability is equally applicable to numeric data and offers the opportunity of data aggregation from different sources into one topical summary. Use of commercial systems in this manner requires careful study of legal aspects and contractual agreements. With regard to the transfer of technology, we are making good use of the following procedure which does not require transfer of the data or models to TIS, and retains full control in the hands of the originators: A computer account is opened for the TIS user community on the remote host computer where the resources are located. The owners of the data and models release, periodically, those versions which they are prepared to share with TIS.





CONCLUSION

I believe we can look forward to the 1980s with excitement. In the past, technology often lagged behind our information requirements. The reverse may soon be true. We probably find it difficult to imagine ourselves using a pocket-portable flat-screen CRT terminal, homed in on a satellite and capable of searching and reading -- wherever we are -- any of the 5.2 million retrospective MARC files of the Library of Congress now being installed on DIALOG, or conducting a molecular structure search on the 4-million-large Structure and Nomenclature Search System, soon to be offered by CAS and CIS. On the home front, the new 9-digit ZIP code of the U.S. Postal Service should be capable, by itself, of delivering mail to anyone in the United States directly by cross-correlation with computer-based address lists and/or social security numbers. Orwell's 1984 is only three years off.

The "Network Nation" is more than the title of a recent book.[78] It marks the beginning of a new decade in which intelligent factual information may become the scarce and costly resource. National and international organizations are trying to come to grips with this situation. CODATA's emphasis is on data in science and technology.[79] UNESCO recognizes basic requirements for factual data in developing countries.[80] Regional information centers in the Far East and Africa are being planned that could transfer the know-how of the post-industrial countries to those anxious to learn, but foresighted enough not to repeat the mistakes of the past. There are also signs of concern, one of which is the abuse of information. We are learning that CIS is bringing online for worldwide access commercially nonconfidential data and production volumes of chemical manufacturing plants in the United States.[81] We are also aware that some countries find it expedient and in their interest to model the economy of the United States with U.S. demographic and time-series data on U.S. computers.

In our democracy, free access to information is essential. We should not be remiss in using it ourselves, first!

May God grant us the wisdom to know how to proceed.[82]

ACKNOWLEDGEMENT

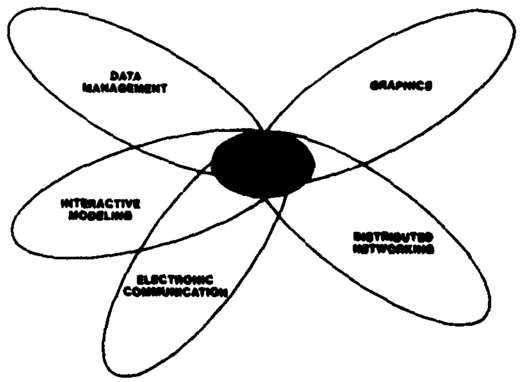
Robert Tap, Director of the Information Division of the Transportation Systems Center for the Department of Transportation (DOT-TSC), specified in 1975 the functional requirements for a transaction controller that could serve as an automated gateway for DOT system analysts. This resulted in the Monitor of Distributed Data Systems (MODDS) and laid the foundation for the work later continued by the Department of Energy.[83]

The Technical Information System (TIS) owes its beginning to the foresight and technical judgement of George F. Pezdirtz, then Director of the DOE Division of Energy Storage. He recognized years in advance that information science and communications technology were on converging paths, and that this merger could be directed toward faster and better access to information communications among administrators and R&D personnel of an important national program, and result in increased productivity. This organization (Office of Energy Systems R&D) has been the prime sponsor of TIS.

The staff of Control Data Corporation and of Schriebman Consulting have contributed unique expertise and talents to the project.

Lawrence G. O'Connell, Manager of the Transportation Systems Research program at LLNL, fostered the development of the Technology Information System. It is used as a practical and useful tool for cross-country information management in support of electric vehicle research.

This report was expertly typed and designed by our Word Processor Specialist Ellen Gerhard, with assistance from Carol Addison, and Donna Scacutto.



REFERENCES

1. The Conference Board: Information Technology - Some Critical Implications for Decision Makers. Published periodically by the Conference Board, Inc., 845 Third Avenue, New York, NY, 10022, 1972.
2. IBM, Information Systems Planning, 5-day Seminars given on the East and West coast by the IBM Systems Science Institute, International Business Machines Corporation, Data Processing Division, 1133 Westchester Avenue, White Plains, NY 10604; IBM World Trade Corporation, 360 Hamilton Avenue, White Plains, NY 10601.
3. Heller, S. R., Economics of Online Data Dissemination, 7th International CODATA Conference, Kyoto, Japan, 1980.
4. Abelson, P. H., Use of Data in Basic and Applied Science, Keynote address, 7th International CODATA Conference, Kyoto, Japan, 1980.
5. The present major computer and communication facilities at Lawrence Livermore National Laboratory, located 45 minutes East of San Francisco, on a 2 mile square area are:

Computers:	4 -	CDC-7600
	4 -	Cray-1
		Numerous mini-computers
Communications:		MFENET
		SACNET
	2 -	Satellite Stations (WESTAR-2 & 3)
	4 -	ARPANET
	16 -	TELENET/TYMNET
	360 -	Dedicated lines
	7000 -	Telephones
Word Processors:	160 -	Word Processors (WP) CPUs
	650 -	WP Work Stations

6. Blumberg, Stanley A., Energy and Conflict. The Life and Times of Edward Teller, G. P. Putnam's Sons, New York, 1976, pp. 95-96.
7. Shimanouchi, T., Two Projects on the Evaluation and Compilation of Scientific Data in Japan, Proceedings of the Fifth Biennial International CODATA Conference, Boulder, Colorado, June 28-July 1, 1976.
8. A Report of the Science Council, A New Plan for a Science Information System in Japan, The Ministry of Education, Science and Culture, January, 1980.
9. Matthoefer, H., The Third Data Processing Program of the Federal Republic of Germany, for 1976-79, published by the Directorate for Research and Technology, Bonn, Bonn University Press, ISBN-3-88135-007-1, 1976.
10. Euronet DIANE Information, Jean Monnet Building, B4 009 CEC, Luxembourg, (Grand Duchy).
11. Committee on Data Needs, Numerical Data Advisory Board, Assembly of Mathematical and Physical Sciences, National Academy Of Sciences, National Needs for Critically Evaluated Physical and Chemical Data, 1978.
12. *Personal communications with Dr. George Michael, Lawrence Livermore National Laboratory, 1981.*
13. Michael, G. A., Mass Archival Storage: Some Trends, Needs and Plans at DOE Laboratories, Lawrence Livermore Laboratory, UCRL-82354 Rev. 1, 1979.
14. Michael, G. A., Archival Storage Requirements, Lawrence Livermore Laboratory, UCRL-82513, 1979.
15. Buzbee, B. L., Worlton, W. J., Michael, G., and Rodrigue, G., DOE Research in Utilization of High-Performance Computers, Los Alamos National Laboratory and Lawrence Livermore National Laboratory, LA-8609-MS, 1980.
16. Landau, R. N., Wanger, J., and Berger, M. C., Directory of Online Databases (Vol. 1, No. 3) Cuadra Associates, Inc., 1523 Sixth Street, Suite 12, Santa Monica, CA 90401, 1980.
17. Williams, M. E., Lannom, L., O'Donnell, R. and Barth, A. N.. Computer-Readable Data Bases, A Directory and Sourcebook. American Society for Information Science (ASIS), 1979.
18. Learned Information, Eusidic Database Guide 1980, Besselsleigh Road, Abingdon Oxford OX13 6EF, England, ISBN 0 904933 20 2.
19. Westbrook, J. H. and Desai, J. D., Data Sources For Materials Scientists and Engineers, General Electric, Schenectady, New York, Report No. 78CRD046, 1978.
    Refer also to: Kruzas, A. T., Encyclopedia of Information Systems and Services, Third Edition. Gale Research Co., Book Tower, Detroit, Michigan, 48226, 1978.
20. Fry, J. P., and Sibley, E. H., The Development of Data-Base Technology, ACM Computing Surveys, Special Issue: Data Base Management Systems, Vol. 8, No. 1, 1976.
21. Hampel, V. E., Database Management Systems for Numeric/Structured Data. 7th International CODATA Conference, Kyoto, Japan, October 1980. UCRL-84594 & Proceedings of Conference.
22. Auerbach Editorial Staff, Data Base Management. Auerbach Publishers Inc., 6560 N. Park Drive, Pennsauken, NJ 08109, 1980.
23. Datapro, A Buyer's Guide to Data Base Management Systems. Datapro Feature Report, Delran, New Jersey, 1980.
24. CODASYL Systems Committee. Selection and Acquisition of Data Base Management Systems. New York, NY: Association for Computing Machinery (ACM); 252 p. ACM Order Dept., P. O. Box 12105, Church Street Station, New York, NY 10249, 1976.
25. CODASYL Systems Committee, Feature Analysis of Generalized Data Base Management Systems, CODASYL Systems Committee Technical Report, May 1977.
26. Seymour, J., Fife, D., Deutsch, D., and Sockut, G., Architectural Considerations For Federal Database Standards, Institute for Computer Sciences and Technology, National Bureau of Standards, Proceedings of 18th IEEE Computer Society International Conference, San Francisco, CA, 1979.
27. Tubbs, N.,. Generalized Data Management Systems and Scientific Information, OECD Nuclear Energy Agency, U.S. Dept. of Energy, U.S. National Bureau of Standards, 38 bd. Suchet, 75016 Paris, France. CONF—771030, 1978.
28. NASA, Engineering and Scientific Data Management. Proceedings of a Conference held at Langley Research Center, Hampton, VA, May 18-19, 1978. ICASE. NASA Conference Publication 2055, 1978.
29. Klingbiel, P. H., A Technique For Machine-Aided Indexing, Defense Supply Agency, Defense Documentation Center, Inform. Stor. Retr. Vol. 9, pp. 477-494. Pergamon Press, 1973.
30. Klingbiel, P. H. and Rinker, C. C., Evaluation of Machine-Aided Indexing, Defense Documentation Center, Cameron Station, Information Processing & Management, Vol. 12, pp. 351-366. Pergamon Press, 1976.

31. Lerner, R. G., Feinman, R., and Lieblich, I., Data-Descriptive Records in the Physical Sciences, American Institute of Physics, 335 East 45th Street, New York, NY 10017, AIP-76-1, 1976.
32. Hampel, V. E., Wiley, R. A., Integrating Text and Data for Energy Research, Lawrence Livermore National Laboratory and Los Alamos Scientific Laboratory, UCRL-76991; also published in the Proceedings of the ERDA-wide Conference on Computer Support of Environmental Science and Analysis, Albuquerque, NM CONF-750706, July 9-11, 1975.
33. Williams, M. E., Annual Review of Information Science and Technology. American Society for Information Science (ASIS), 1979.
34. Organized each year by Online Review, *The International Journal of Online & Teletext Information Systems,* Online Information, Learned Information, Oxford and New York.
35. Annual Conferences on Very Large Databases.
36. Williams, M. E., Research Toward a Transparent Information System, Information Retrieval Research Laboratory, Proceedings of Perspectives on Scientific and Technical Information, Oak Ridge, Tenn, ORNL-5749, 1980.
37. Killeen, J. and staff, The National MFE Computer Center: Increased Capability in 1978, Energy and Technology Review, Lawrence Livermore National Laboratory, September 1978.
38. Hampel, V. E., and Henry, E., The Time for Atomic and Molecular Databases is Now!, Lawrence Livermore National Laboratory, UCRL-79286, 1977.
39. Hilsenrath, J., *Summary of On-Line or Interactive Physico-Chemical Numerical Data Systems,* Office of Standard Reference Data, National Bureau of Standards, Washington, D.C., NBS Technical Note 1022, 1980.
40. Hall, D., Scherrer, D., Sventek, J., A Virtual Operating System, Communications of the ACM, September 1980.
41. Donnelley, J. E., Managing Domains in a Network Operating System, Lawrence Livermore National Laboratory, UCRL-85562, Proc. of the International Conference on Local Networks & Distributed Office Systems, London, May 11-13, 1981.
42. General Purpose Computer Networks and Resource Sharing in ERDA, Volume 1, 2, 3, Summary, Conclusions and Recommendations, Working Group on Computer Networking, Division of Basic Energy Sciences, ERDA, July 15, 1977.
43. Proceedings of the Berkeley Conferences on Distributed Data Management, and Computer Networks, CONF-790834, 1979 and 1980.
44. American Society for Testing and Materials, Standard for Metric Practice, 1916 Race Street, Philadelphia, PA 19103.
45. Hampel, V. E. and Ries, D., *Requirements for the Design of a Scientific Data Base Management System,* Lawrence Livermore Laboratory, UCRL-80171, Rev. 1, Proceedings on Generalized Data Management and Scientific Information, sponsored by the OECD Nuclear Energy Agency, U.S. DOE, & National Bureau of Standards, Lawrence Berkeley Laboratory, CONF-771030, 1977.
46. David, A. and editor, Engineer's Needs for Scientific Technical Information, World Federation of Engineering Organizations, Committee on Engineering Information, Paris, 1979.
47. Selander, S. E., A Syntax for Machine-Readable Description of Numeric Data, Mitre Corporation, 1979.
48. Winter, R., Lozano-Perez, T., Eastlake III, D. E., and Rosenberg, R. L., Overview of the Chemical Substances Information Network, Computer Corporation of America, 575 Technology Square Cambridge, MA, CCA-81-03, 1981.
49. Lide, D., *Critical Data for Critical Needs, Science,* Vol. 212, No. 4501, 1981.
50. The National Commission on New Technological Uses of Copyrighted Works, Final Report, Library of Congress, 1978.
51. Bush, G. P., Dreyfuss, R. H., and editors, Technology and Copyright: Sources and Materials, 2nd Edition, Lomond Publications, Inc., Airy, MD, ISBN 0-912338-17-2, 1979.
52. Uhlig, R. P., Farber, D. J., Bair, J. H., The Office of the Future, Communications and Computers, North-Holland Publishing Company, 1979.
Refer also to: National Research Council, Board on Telecommunications-Computer Applications, Toward the Electronic Office, Washington, DC, Guidance on Requirements Analysis for Office Automation Systems, NBS Speciald Publication 500-72., U.S. Department of Commerce, 1980.
53. Woolfe, R., *VIDEOTEX - The New Television/Telephone Information Services,* Heyden, 1980.
Refer also to: Marsh, F. E., VIDEODISC Technology, DOE Computer Operations Branch, Technical Information Center, CONF-810322-1, 1981.
Refer also to: Hirsch, VIDEOTEX in the U.S., Computer World, Washington, DC Bureau.
NOTE: Viewdata and Online Systems and The Potential Impact of Viewdata and Online Retrieval Systems 1980-2000 are some of the first courses offered in this exciting field by H. Leonard Fisher from Lawrence Livermore National Laboratory, as the instructor, at Golden Gate University, San Francisco, CA.
54. Kurzweil, Computer Products, Inc., 33 Cambridge Parkway, Cambridge, MA 02142.
55. Lempel, A., Crytology in Transition, Computing Surveys, Vol. 11, No. 4, 1979.
Refer also to: National Bureau of Standards, Federal Information Processing Standards, Publication 46, 1977.
Refer also to: Diffie, W. and Hellman, M. E., New Directions in Cryptography, IEEE Trans. Inform. Theory IT-22, pp. 644-654, 1976.
56. Hampel, V., McGrogan, S. K., Gallo, L. E., and Swanson, J. E., THE LLL "META-MACHINE"--A Flexible, Extensible, and Practical Technique for Interactive Data Management, Modeling, and Distributed Networking, Lawrence Livermore National Laboratory, Livermore, CA, UCRL-83064, May 1979.
Refer also to: Hampel, V. E., et al, The META-MACHINE - a New Way of Making a Mini-Computer Do a Maxi-Job in Data Management, Lawrence Livermore National Laboratory, UCRL-81170, Presentation to the CUBE Symposium, October 4-5, Los Alamos, NM, 1978.
57. Woodfill, J. et al., INGRES, Version 6.2, Reference Manual, Electronics Research Laboratory, College of Engineering, University of California, Berkeley 94720, 1979.
Refer also to: Epstein, R., Creating and Maintaining a Database Using INGRES, NCB/ERL M77/71, 1977, and to the review article on other current relational systems by Rita Shoor in the February 9th, 1981, issue of "Computerworld."
Refer also to: Schneidawin, J., Users of Bell's UNIX Seen Spending $9.58 Billion by 1986, Management Information Systems, p. 17, July 22, 1981.
58. The Intelligent Database Machine (IDM-500), Britton Lee, Inc., Cupertino, CA.
Refer also to: The excellent review article on other database machines by Vincent C. Rauzino in the December 29, 1980/January 5, 1981 issue of "Computerworld."

59. Koponen, B. L., Wilcox, T. P., and Hampel, V. E., Nuclear Criticality Experiments from 1943 to 1978, An Annotated Bibliography in Three Volumes, Lawrence Livermore National Laboratory, UCRL-52769, 1979.

60. American Association for the Advancement of Science, SCIENCE, Vol. 211, No. 4483, February 13, 1981.
Refer also to: Lawrence Berkeley Laboratory, NEWS MAGAZINE, Vol. 5, No. 2, Summer, 1980.
One of the best known products developed by the LLNL Graphics Group is the portable, generalized, color graphics program: Rowe, J. Keller, P., and O'Hair, K., GRAFCORE Reference Manual, Lawrence Livermore National Laboratory, LCSD-434, 1980.

61. Vickers, D. L., Smith, G. S., and Levine, S. R., Lawrence Livermore National Laboratory, and L. G. Cross, The Multiplex Company, San Francisco, California, Two Unusual Three-Dimensional Displays, Trans. Am. Crystallographic Assn., Vol. 12, 1976.

62. Vickers, D. L., The Integral Hologram as a Scientific Tool, UCID-17035 Lawrence Livermore National Laboratory, California, 1976.

63. Jeong, T. H. and Snyder, H., SID Int'l. Symp., Digest of Technical Papers 5, 1974.

64. Sutherland, I. E., Proc. Fall Joint Computer Conference 33, 757, 1968.

65. Hampel, V. E., Gallo, L. E., Kawin, R. A., Kopytoff, V., McGrogan, S., O'Connell, L. G., Pavel, G., Schriebman, J. A., Swanson, J. E., Editor, Proceedings of the First DOE/LLL Workshop on the Technical Management Information System (TMIS), CONF-791258, 1979.

66. Hampel, V. E., Gallo, L. E., Kawin, R. A., Kopytoff, V., Mallon, B., McGrogan, S., Pavel, G., Rabe, W. G., Schriebman, J. A., Scott, W. S., and Swanson, J. E., User's Manual for the Technology Information System, 2nd Edition, M-0112, 1981.
Refer also to: Hampel, V. E., Decision Making with Interactive Access to Integrated Administrative and Technological Databases, Lawrence Livermore National Laboratory, UCRL-80353, Proceedings of the ERDA/AESOP XVII Conference, Boston, MA, September 13-15, 1977.

67. The Transportation System Research program at LLNL conducts theoretical and experimental research for future electric vehicles, the innovative aluminum-air battery, roadway powered electric vehicles, route analysis for the U.S. Postal Service., and transportation-related R&D in general. The computational analysis and performance predictions are made on the Technology Information System. Some are cited here to convey the complexity of calculational-graphics, and the extent of information & data management carried out on the TIS intelligent gateway computer:
   a) O'Connell, L. G., et al., Energy-Storage Systems for Automobile Propulsion: Final Report, Volumes 1 - 4, Lawrence Livermore National Laboratory, Livermore, CA, UCRL-53053-80, 1980.
   b) Behrin, E. and Cooper, J. F., Aluminum-Air Batteries for Vehicles, Lawrence Livermore National Laboratory, Livermore, CA, UCRL-85605, 1981.
   c) Walter, C. E. and Salisbury, J. D., Inductive Energy Transfer in Transportation, Lawrence Livermore National Laboratory, Livermore, CA, UCRL-85141, 1980.
   d) O'Connell, L. G., A Comparison of Energy-Storage Devices for Use in Future Automobiles, Lawrence Livermore National Laboratory, Livermore, CA, UCRL-85654, 1981.

68. Simmons, W. W., The NOVA Laser Fusion Facility, Lawrence Livermore National Laboratory, Energy and Technology Review, 1980.

69. Janz, G. J., Allen, C. B., Downey, Jr., J. R., and Tomkins, R. P. T., Physical Properties Data Compilations Relevant to Energy Storage--Molten Salts: Eutectic Data, Molten Salts Data Center, Cogswell Laboratory, Rensselaer Polytechnic Institute, Troy, NY, NSRDS-NBS 61, Part 1, March 1978.

70. Whisler, L. J., Behrin, E., and Salisbury, J. D., Aluminum-Air Vehicle Simulation Programs User's Manual and Documentation, Lawrence Livermore National Laboratory, Livermore, CA, M-113, 1980.

71. Kong, M. K. and Walter, C. E., Route Profile Analysis System Report Covering Task A - CARA Model Verification, Lawrence Livermore National Laboratory, Livermore, CA, UCID 19027, 1981.

72. Chapman, P., Hybrid Vehicle Simulation Program (ELVEC Version 6.0); Volume I: Program Documentation, Jet Propulsion Laboratory, 5030-247, Rev. D, 1981.

73. Mangold, D. C., Lippmann, M. J., and Bodvarsson, G. S., Computer Program for the Numerical Solution of the Heat and Mass Flow Equations for a Liquid Saturated Medium - User's Manual, Lawrence Berkeley Laboratory, LBL-10909, 1980.

74. Hampel, V. E. and Schriebman, J. A., A Personal System for Electronic Mail, Lawrence Livermore National Laboratory, Manual M-115., November, 1980.
Refer also to: Schriebman, J. A. and Swanson, J. E., Crosslinking of Computer Terminals for Nationwide Tutorial Instruction, Lawrence Livermore National Laboratory, UCRL-52964, November, 1980.
Refer also to: Information Systems: Voice & Data's Mailbox Uses Any Phone, Management Information Systems, July 15, 1981.

75. Rosenthal, R. and Lucas, B. D., The Design and Implementation of the National Bureau of Standards' Network Access Machine (NAM), Institute for Computer Sciences and Technology, National Bureau of Standards, NBS Special Publication 500-35, 1978.

76. Rosenthal, R., A Review of Network Access Techniques with Case Study: The Network Access Machine, National Bureau of Standards, NBS Technical Note 917, July 1976.

77. Dubois, J. E., Analysis of Data Bank Implementation and Computer Aided Design in Chemistry, Proceedings of the Fifth Biennial International CODATA Conference, Boulder, CO, 1976.

78. Hiltz, S. R., Turoff, M., The Network Nation, Human Communication via Computer, Addison Wesley Publishing Company, 1978.
Refer also to: Fedida, S., Malik, R., Viewdata Revolution, John Wiley & Sons, 1979.

79. Watson, D. G., Rossmassler, S.A., Data Handling for Science and Technology, An Overview and Sourcebook, CODATA & UNESCO sponsored, North-Holland Publishing Co., 1980.
A corollary publication on Database Management in Science and Technology is in preparation.

80. UNISIST Newsletter, General Information Programme, Vol. 8, No. 4, 1980, and Vol. 9, No. 1, 1981.
Refer also to: Study on an International System Relating to New and Renewable Energy Sources, UNESCO, General Conference, 21 C/INF.10, Belgrade, 1980; Report on the ad hoc Expert Group on Information Flows, A/CONF.100, January 1981.

81. New Chemical Database has Non-Secret Plant Information, Announcement by NIH/EPA, July 1981.

82. The Bible, The Epistle of James, Chapter 1, Verse 5.

83. Birss, E. W., Donnelley, J. E., and Yeh, J. W., A Monitor of Distributed Data Systems (MODDS): Part-1 and 2, Digest of Functional Specifications for the Department of Transportation, UCID-17314, Lawrence Livermore National Laboratory, November 15, 1976.

USER REQUIREMENTS IN HANDLING ONLINE SYSTEMS AND NETWORKS

Aud Lamvik
Head of Documentation Department
University of Trondheim
Norwegian Institute of Technology
The Library
N-7034 Trondheim-NTH
Norway

Online systems and networks used by Norwegian searchers are outlined. The Nordic information network SCANNET is described. User problems and needs are presented, in the light of the work in online user groups.

Online information systems are young, only a little more than a decade. This new information technique has developed strongly since the very beginning, and user requirements in handling such systems have proved to be valuable factors in the development of online services as well as networks. As users of online databases become more experienced and better aware of the different search possibilities, they make recommendations which should affect the further development of the online services, and I think they will be influential even to a larger measure during the eighties, in forming the future online information services.

DEFINITIONS

First, let me explain who are the users. They are primarily the searchers of the online databases, in other words the intermediaries, brokers, information officers, those who are formulating the search strategies, selecting the systems and choosing the files, using the terminals and keeping up with the news and changes of the services.

But there are also the end-users, those who are not actually doing that preparatory work, but are the real users of the information in their work.

Thirdly there are the non-users, users who probably are not aware of the online services and are potential future users.

All groups have their particular requirements, and it is obvious that the system and network designers must take such requirements into consideration in order to make their online services as attractive to the users as possible. There will be a greater struggle in attracting users to the systems as competition grows.

The online systems are in this connection the systems we are using for information retrieval, with bibliographic and referral as well as numeric and fact-oriented databases.

TODAY'S SERVICES

My reflections upon the user needs are based on the online databases, systems and communication networks we are using in Norway today, especially in the fields of science and technology. Let me start with an overview of the Norwegian situation, where my basis is the systems and networks used by the library of the Norwegian Institute of Technology. Our library documentation services are offered to the staff and students of our university as well as to our home industry, research institutions and to the public, in keeping with our library's assumed duties as central library of technology for our entire country.

We started the online searching in 1973 with MEDLINE, by direct long distance calls to the database in Sweden, after some years of batch SDI experiences. In 1975/76 we started as online users of the Lockheed, SDC and ESA-IRS systems. In 1978 we installed a 1200 baud terminal, and that higher speed is now used in the online searching when possible.

The online information services and networks we are using today are:

DIALOG	TYMNET	Amsterdam
SDC	"	"
NLM	"	"
ESA-IRS	ESANET	Copenhagen, Stockholm
SCANNET	NORPAK	Oslo
"	UNINETT	University of Trondheim
TECHNOTEC	CYBERNET	Oslo
BLAISE	Direct call	

In addition we are using some other online databases, e.g. library catalogs as our own BIBSYS. This means several different passwords to networks, systems and d ses.

The users of our documentation services are primarily interested in bibliographic data. We are concerned about the growing number of numeric databanks, but the library is well equipped with reference literature and data books in all fields of interest, so there are

not many questions about searching online numeric databanks. There is a growing demand for market/business information, but the bibliographic information in science and technology is still most asked for.

From Norway we get connected to the US information systems by calling the Amsterdam TYMNET node. Changes in that procedure will be made in the near future, when the Norwegian NORPAK node will be connected to US information networks.

Norwegian online users have no access to EURONET, but we are waiting for it, for instance through the Scandinavian network which will be described.

SCANNET NETWORK AND DATABASES

SCANNET is the Scandinavian contribution to networks and databases. It was established by NORDFORSK, the Nordic Co-operative Organization for Applied Research, in 1976 as an experimental network to make machine-readable databases and -banks accessible online within the Scandinavian countries. It went into full-scale operation in 1976 (1). Plans to interface SCANNET with EURONET and other networks were made.

In 1979, when SCANNET was reorganized into a foundation, the Nordic PTTs decided to make SCANNET a "semi"-public network. The telephone lines were no longer leased by SCANNET, but the searching procedures for the users remained the same. Today the Scandinavian network is a packet-switching service in the telephone network (2). Since the PTTs took over the responsibilities for the communication part, we no longer have the old SCANNET. The network consists of 2 nodes, Databas 300/1200 or TELEPAK in Stockholm, and NORPAK in Oslo. Multiplexors in Copenhagen and Helsinki are connected to TELEPAK, which is also connected to information networks in North America. The previous node computers in SCANNET now have a function as front-end processors for the various host computers connected. The function of SCANNET in the network can probably be compared to DIANE in EURONET/DIANE.

A universities' network in Norway, UNINETT, has been developed and connected to NORPAK. It can, therefore, be used for searching the SCANNET databases. The connection from TELEPAK to the networks in the USA is not open to Norwegian users, but as mentioned earlier, there are plans to connect NORPAK to networks in the USA.

The number of SCANNET databases offered is constantly increasing. Databases are (spring 1981):

ALIS	- Technical literature	
BVR	- Swedish building market - products and suppliers	
BYGGDOK	- Community planning, architecture etc.	
CHEM	- Factual bases on crystallography and mass spectrometry	
CIS/ILO	- Occupational safety and health	
EPOS	- Profile handling system for VIRA, SDI-service	
EXTEMPLO	- Electronic journal within library and information sciences in the Nordic countries	
FINP	- Finnish economics and business periodicals index	
INIS	- Nuclear energy & technology	
IRRD	- International road research documentation	
MARKNADSBANK	- Swedish market information bank	
MECHEN	- Mechanical engineering	
MEDIC	- Finnish medical literature	
MEDLARS	- Medicine, biomedicine. Several databases:	
	MEDLINE	
	CANCERLINE	
	CHEMLINE	
	RTECS	
POLYDOC	- Several databases:	
	AID	- Artikkel-Indeks Database from journals in technology and management
	AIPG	- AJOUR and product guides in technical fields
	SDI	- Nordic literature in library, documentation and information
	ETABLERING	- Norwegian industrial establishments abroad
	EKSPORT-INDEKS	- Export markets
	FoU-INDEKS	- Norwegian R&D Index
	OLJE-INDEKS	- Oil Index
	SHIP ABSTRACTS	- Ship technology, ocean engineering
	STANDARDS	- Norwegian standards
RECODEX	- Locating index for interdisciplinary scientific reports in Scandinavia	
SCANP	- Scandinavian economic and business periodicals index	
SCIMP	- European periodicals in business, management etc.	
SERIX	- Swedish environmental research index	
VANYTT	- Environmental technology	

Some test databases are also in the system.

As you can see, about 30 databases are available. Some of them are well known international ones, while others are small, specific, Scandinavian databases and -banks, most of them searchable in a Scandinavian language which of course is a barrier for international

use. Yet you may find information in them which is not available in other databases, and I am sure the interest for the SCANNET databases will be growing also outside Scandinavia.

NORDINFO, the Nordic Council for Scientific Information and Research Libraries, has now taken over the financial responsibilities for SCANNET from NORDFORSK and is also responsible for value additions and the building up of databases in the information network (3). In supporting databases, priority has been given to those which have a Nordic connection and contain material that is not available via the large international information systems. Many of the existing databases in SCANNET have been supported by NORDINFO, and further development of bibliographic and factual databases will surely result in several new SCANNET databases.

ONLINE USER GROUPS

It seemed in the first years of the "online age" that people were mainly discussing how computers could be used in storage and dissemination of information. The user needs were seldom elucidated. But the many information systems and networks soon gave the online searchers various practical problems in their work. During the last years several studies on user needs have been published.

In 1976-78 online user groups started their activities to focus their attention on the online user problems and needs, and to extend and spread the knowledge of information systems and networks. At the First International On-Line Information Meeting, 1977 in London, representatives of online user groups met informally to form a European group, which is now a working group of the European Association of Information Services, the EUSIDIC Working Group "National User Groups". One of the aims is to represent the interests of the users, and the group has discussed many problems and made several recommendations.

NORDINFO took the first steps to elucidate the Nordic online user problems and needs by bringing together representatives of the Nordic online users in 1978. This meeting resulted in a report to NORDINFO on the situation (4) and further cooperation between Scandinavian user groups, who have been meeting for the last few years to discuss problems in common.

In the light of the activities in the user groups, and from the experience I have as online user, especially in the fields of science and technology, I call attention to the following user problems and requirements:

- Passwords, logging in procedures, telecommunication problems
- Choice of systems and files to get the best information at the lowest price
- Search capability, knowledge of database structures, indexing practice
- Availability of databases, systems and networks, exclusive rights
- Pricing policy, annual fees versus prices for actual use, type charges
- Standardization
- Back-up services, easy access to the primary documents, online ordering
- Administration of services, statistics, updating of user aids, marketing
- Education and training, experienced and new users, information officers and end-users (and non-users)
- Terminals

The passwords are often complicated even for an experienced user. Although the searchers are getting used to the different ways of connecting to the systems and databases, the "logging in" procedures, a uniform mode of operation with simple passwords is urgently required, especially for small users. Network problems cause extra costs. It seems that there are more often communication breakdowns than there are problems with the host computers, systems or search programs. Reliable networks are necessary for efficient searching. In Norway the telecommunication costs are more than the computer and database costs when using the US information systems.

It is important to have access to various systems, to be able to choose. When one system is out of function, another may be used if a search is urgent. Different systems may offer the same databases, but users should be aware that there are sometimes differences in the way they can be searched. Search commands differ from one system to another, and various features are made searchable. A database can be split into sub-files in one system, while kept complete in others. Prices can vary between hosts, and computer responses may differ, which also affects the search costs.

Search capability, knowledge of database structures and indexing practice are important factors in the online systems to get the best information at the lowest price. Experience from printed abstracts journals is valuable, and here the libraries have an advantage in their large abstracts journals holdings.

Access to some databases and networks is currently restricted in various ways, e.g. geographical. European users have no access to some of the DIALOG databases, Norwegian users have no access to EURONET etc. The database exclusive rights probably originates from the database suppliers, but the goal must be equal access for all users to publically available databases.

The time differences between countries and continents have been a problem for database access during users working days. The service schedule for the USA information systems seems now more satisfactory for European users, but there are still systems where the time sche-

dule may be a problem, e.g. in SCANNET.

During the last few years there have been changes in policy and tariffs adopted by vendors and database suppliers. Paying for the actual use of a database is preferable. Annual or monthly taxes are a drawback being often too expensive for small users. When paying for the actual use, it is easier to invoice the end-user.

Type charges which have been brought on during the last few years, are an unjustified burden on the interactive search process. It is important to study references while doing an online search to make it as efficient as possible, and some type formats without charges are necessary. Fortunately such formats seem to be common now in information systems.

There is an increasing demand for standardization, especially at the user-system interface, but it should cover databases as well. It is difficult, especially for small users, to be effective when every system has its own special routines. Research on a common command language is in progress, but one should have in mind that database structure and specialities always have to be important factors for effective searching.

A standardized output is preferable. The variations in printed references cause problems for end-users. References should be easy to read and understand.

It would be desirable to have a standardized invoicing practice, which is one of the points the EUSIDIC user group has been concerned with.

When operating an online service it is important to offer a satisfactory back-up literature service to the users. The online searching is only a part of the information retrieval process, and the full text of a document must be made available to the user.
The large libraries have here an advantage in their collections and their national and international interlending partners. Studies are being undertaken concerning UAP, Universal Availability of Publications, - the achievement of the widest possible availability of published material to intending users, wherever and whenever they need the information.

Online ordering is a rather new facility introduced by several online vendors. The service is often more expensive than ordering from a library, but it may be an efficient way to obtain the full text documents at a reasonable length of time. Document suppliers in the online ordering facilities may be database producers, libraries or information brokers. It seems that database suppliers and libraries here will be giving the best service, but until now few studies of the service have been published.

Education and training is important on all levels, for new and experienced searchers. System-independent education is desirable, in addition to the training courses sponsored by database producers and online vendors. There have been prepared many user aids and workbooks, also for self-instruction, which I find very useful, because going to a lot of different training courses is time consuming and expensive. I suppose that education and training will be further developed during the years to come. The user groups are very much concerned in this matter.

Choosing a terminal for online searching is largely a matter of personal requirements. Many variations are possible - display, printer, combination of both, portable, integrated acoustic coupler, switchable speed or parity, small memory, programmable, etc.
I prefer a video display terminal and printer, with switchable speed or parity, and additional monitors for demonstrations and teaching. A memory is useful, and new aids to users are constantly being developed, for instance microprocessor-assisted terminals for more efficient searching.

Much has been done during the last few years to make online services attractive to the users. Many system and network design problems have been solved. Vendors and producers have realized that information to users, user aids and guidances, are important, and that information about services needs frequent updating. The users receive more information today than some years ago. The administration of online services is manifold, and users are very much concerned about it. A better understanding of the user, the information officer as well as the end-user, the psychology of man-computer interaction and the human factors affecting the quality of this interaction, seems to be important. Some investigations have been made in recent years (5).

Information should be equally available to whomsoever seeks it. The value of information is however not realized before it is actually used by the end-user, the research scientist or decision maker. Our knowledge is still highly insufficient in this matter. I want to conclude with some words by Y.S. Touloukian, which I find important: "While the scientific and technical community is being smothered by an overwhelming document birth rate, it starves from a lack of evaluated information" (6). That is a problem also for the online user in his effort to find the best information.

REFERENCES

(1) Abrahamsson, S.: SCANNET and its relations to the PTTs. Proceedings of the EUSIDIC Conference, Berlin (West), 9-11 November 1977, p.69-72.

(2) SCANNET Today. Information on available data bases. Stockholm, Winter 1980-81.

(3) NORDINFO. The Nordic Council for Scientific Information and Research Libraries. Esbo, (1981) p.10-13.

(4) Nordiska Online-Användare - Ad hoc-möte december 1978. NORDINFO-Nytt, Vol. 2, No. 2, 1979, p.36-44.

(5) Baker, C.A. and Eason, K.D.: An observational study of man-computer interaction using an online bibliographic information retrieval system. Online Review, Vol. 5, No. 2, 1981, p.121-132.

(6) CINDAS, Center for Information and Numerical Data Analysis and Synthesis. Twenty Fourth Annual Report - 1980.

Optical Disc Technology and Its Implications for *Information Storage and Retrieval in the Eighties*

Charles M. Goldstein
Lister Hill National Center for Biomedical Communications
National Library of Medicine
Bethesda, Maryland 20209
USA

SUMMARY

Optical disc technology includes the storage and retrieval of random access graphics on videodiscs as well as the storage of digital information. Optical digital discs for computer mass storage are currently under development by many firms. In addition, efforts are underway to allow the encoding of digital information on videodiscs. The latter is desireable as an inexpensive publication medium for machine readable data as well as a means of obtaining both video and digital information on one disc. Potential applications of this technology include inexpensive online storage, random access graphics to complement online information systems, hybrid network architectures, office automation systems, and archival storage.

II. INTRODUCTION*

The optical videodisc and its counterpart, the optical digital disc, both hold great promise for information storage and retrieval (IS&R). The videodisc is already finding applications in the delivery of graphics to complement online interactive information systems as discussed in Section IV.A. In addition, the application of optical discs to digital storage represents one of the most promising developments in computer mass storage technology in recent years. While this new technology is undergoing a rather long gestation period, its potential for IS&R should be realized in the 80's.

This paper shall first review the techniques for storing digital information on optical discs. Greater detail may be found in [1]. Subsequent sections shall address the potential impact on IS&R.

III. OPTICAL VIDEODISC

A. Recording Technology

There is considerable activity, internationally, in the development of various videodisc recording technologies. This paper is limited to optical disc technologies because of their special interest to IS&R. Even, here, there exist different approaches. Fortunately, two firms that have pioneered in bringing the videodisc technology to the marketplace, MCA (now DiscoVison Associates or DVA) and Magnavox-Philips, have agreed on a common disc format. This format contains information coded in the form of microscopic "pits" in a reflective information surface. The information surface is covered by a transparent plastic layer. Playback is performed by focusing a laser light beam on the information surface and monitoring the reflected light pulses. No light is reflected from the pits. The reflected light is, in reality, a frequency modulated signal resulting from the differing lengths of the pits and the "land" between them. Thus, the encoding is not just a series of 0's and 1's as is so customarily the case for computer storage. There may be as many as 14 billion (14,000,000,000) pits per side of a one hour (2 sides) videodisc with spacings between concentric tracks of about 1 micron (39 millionths of an inch). With these dimensions, dust particles of just a few microns in diameter at the information surface could obliterate many signal elements. The protective transparent surface serves to keep such dust away from the information surface and outside of the focal plane of the "optical stylus", thus minimizing loss of information.

There do exist optical recording technologies other than those utilizing reflecting surfaces and pits. These include both photographic and photochromic processes. A list of such processes is given in [2]. In addition, the Thompson CSF Videodisc System employs a transparent, non-reflecting disc. In this system it is the transmission of light through the disc rather than light reflection that provides the information carrying signal [3].

B. Production of Videodiscs

Optical videodiscs are produced by a mastering and replication process [4]. In the DVA process, material to be mastered is first recorded on, or transferred to, a two inch helical video tape. A glass master disc is created by first exposing the prepared, photosensitive surface with a laser beam modulated by the output of the signals recorded on the two inch helical video tape. The glass master is then photoetched. From the glass master, multiple nickel metal "stampers" are produced by a plating process. These stampers are then utilized in an injection mold process to stamp one side of the disc.

* Portions of this paper originally appeared in [1]. Permission to include them here for completness was kindly granted by Knowledge Industry Publications, Inc.

After the addition of a reflective coating, two sides are joined together to form a single two-sided disc.

The mastering and replication of videodiscs requires controlled clean-room environments and major investments in sophisticated equipment and facilities. It is a publication process. (It is interesting to note that most traditional publication processes utilize some form of "stamping", be it print on paper, or music on records.)

Production capabilities for the DVA/Philips formatted videodiscs are presently available through both DVA and SONY. Costs for creating the master may vary between $2,200 to $4,000 or more per side depending on the amount of pre-editing or processing required. For small quantities DVA is quoting replication costs of $10 dollas per disc (two sides). Costs for large quantities, presently determined by negotiation, should drop sharply with the size of the production run. 3M Corporation is to establish mastering facilities for the production of Thompson discs in the U. S.

C. Formats and Players

Videodisc players and disc formats may be differentiated as to Consumer or Industrial (or Institutional) types. The consumer and industrial players differ in the amount of control, intelligence, signal quality and robustness offered. Consumer models sell for around $800, while single unit costs for industrial models are around $3,000. It is the latter that will be of primary interest to the IS&R community. Consumer model players are presently being marketed by Magnavox and Pioneer. Industrial players are being marketed by SONY, DVA, and Thompson CSF. The newly announced RCA videodisc players employ a physical stylus and grooved videodisc records; they are of no potential interest for IS&R since they preclude random access and still or "freeze" frame capabilities.

Industrially formatted discs contain concentric tracks with one video frame per track. Each side of a 30 cm. (12 inch) videodisc contains 54,000 frames in the USA standard video format, or 47,000 frames for the higher resolution European video formats. This one frame per track format is that which allows random access and freeze frame capabilities, both of critical importance for IS&R as discussed later. At 30 frames per second, the USA formatted discs have a video playback time of one half hour per side.

DVA has also introduced an optical disc format for the entertainment market that allows one hour of continuous playback per side, or two hours per disc. This format encodes the frames in a continuous spiral, and, like the RCA videodisc system, is of no interest for IS&R since it, too, precludes random access and freeze frame operation.

IV. DIGITAL STORAGE

While the spur to the development of the optical disc technology was the entertainment market, it was also recognized that this technology could be exploited to store digital information. Two approaches have been identified: one is to encode digital information onto a standard video signal and then obtain the videodiscs via the mastering and replication process previously described. This approach will be discussed in Section B. The desirability of recording digital information <u>on site</u> at the computer was also recognized and a methodology for accomplishing this was initially developed by Philips Laboratories, North American Philips, under contract to the USA Department of Defense's Advanced Research Projects Agency (ARPA). This second approach will be designated "optical digital disk" technology to distinguish it from the digital applications of the "optical videodisc".

A. Optical Digital Discs

1. <u>Philips DRAW Process</u> The Philips optical digital disc technology[5,6] utilizes a specially prepared disc. Two glass or plastic discs are coated on the inside surfaces with a special reflective metal film (usually tellurium), separated by an air space, and hermetically sealed. The optical digital disc is mounted in a recorder/player connected to the computer. A relatively high-powered laser beam is focused through the protective surface so that the information is "burned" into the inside information surface. In this recording process, the disc carries its own "clean room" with it. Another advantage is the ability to read the data right after it is recorded. Philips has designated this the "Direct Read After Write" or "DRAW" (trademarked) process. Once the information has been burned into the information surface it cannot be erased and is, consequently, "read-only". In the Philips DRAW information system, data are recorded on 40,000 tracks per side of a 12 inch (30 cm.) disc for a total capacity of 10 billion bits (1 billion characters) per side. Each track of the optical disc contains 32 addressable sectors of 15,200 bits each. Research is underway to extend the data storage to 100 billion bits (10 billion characters) per side. A more detailed discussion of the DRAW process may be found in [6].

2. <u>DREXON DRAW Discs</u> The disc used in the Philips DRAW process, described above, is susceptible to deterioration if air permeates the seals between the two surfaces and oxidizes the metallic film containing the information. A new process, developed by Drexler Technology Corp., [7-9], utilizes extremely small metal spheres and filaments, distributed throughout a polymer. The absorption of the laser light by the metal filaments (silver halide) causes the local reflectivity to change. These micron-wide areas of lower reflectivity correspond to the "pits" in the Philips-type disc. Since the metal particles are dispersed within the disc, however, DREXON discs are impervious to oxidation [10]. As

another advantage, low power diode lasers can be used to both write and read.

In the future, this technology may also be applied to the production of videodiscs, or allow the inexpensive replication of DRAW discs by using a transmissive DREXON master. Copies would be achieved by exposure of light through the master. Today, in small quantities, DREXON 12 inch diameter discs cost $3,500 each.

3. Digital Encoding Since optical digital discs are being developed solely as computer mass storage devices, the information is encoded in a form most suitable for digital recording and requires no compatibility with video signals or formats. The ability to read the data as they are recorded allows the immediate identification of errors. The errors may then be "corrected" by rewriting the data in a new sector and erasing the address of the bad sector from the computer memory. This does not mean that the information can be erased from the disk, but, rather, the knowledge of that sector can be erased from computer memory. Hence, the user information stored on an optical digital disc can have a very high accuracy (low error rate).

4. Applications The optical digital disk, therefore, represents a very high-density, low-cost, archival mass storage medium. Future videodisc storage units may employ jukebox arrangements for the online access of as many as 1,000 discs[6]. Also under consideration are optical disc packs of six double sided discs accessed by 12 separate read heads[6]. Each such disc pack would have a storage capacity of 1 trillion bits or 100 billion characters in the same volume as required today by a magnetic disc pack containing 300 million characters. Optical disc packs will utilize smaller, solid-state laser diodes[11] instead of the bulkier gas discharge lasers, which will allow the dimensions between disks to be much closer.

5. Availability Although announced in 1979, problems have delayed the introduction of commercially available "draw" units. North American Philips announced the availability of prototype DRAW Recorder/Player units for $150,000 and player units alone for $20,000 [12], but then later withdrew the offer. Both RCA and Philips will entertain offers to deliver "DRAW" units but cost must be negotiated.

It is generally known that other firms are developing optical digital disk mass storage systems similar to the Philips system but, as of this writing, no other units have been publicly announced. As noted by Jerome Drexler, President of Drexler Technologies, "...we are currently supplying media to 18 commercial firms...16 are major companies."[13] A review of other firms engaged in the optical disk technology may be found in [13].

B. DIGITAL INFORMATION ON OPTICAL VIDEODISCS

The potential for encoding digital information on the standard NTSC video signal and format was addressed by George Kenney in a 1976 paper [14]. In this paper, he pointed out the economic advantages that could accrue by exploiting the already existing videodisc production facilities for the publication of machine-readable information.

There are, however, many problems to be resolved in the successful exploitation of the videodisc technology for the storage and retrieval of digital information. Paramount among these is the identification and correction of errors. Unlike the DRAW process discussed in the previous section, the mastering and replication process for videodisc production does not allow dynamic error identification and correction. Furthermore, it is known that there are many sources of errors in the production process, including errors in mastering and stamping, and errors due to particulates in the plastics. Errors presently occur in the production of videodiscs for entertainment purposes. These errors (drop-outs), while often visible (drop-outs), are seldom obtrusive to the eye or ear. The sensitivity of digital information to errors is much greater, since the loss or change of a single bit could cause a computer program to malfunction, or an index to return wrong information.

Another major constraint of this recording mode is the need to encode the information into 54,000 frames of basic video (NTSC) format. While a constraint on the one hand, this is also a major advantage; viz., the ability to intermix digital and video information on the same disc. For, if the output of the disc is to be under computer control, then there must be an index to all significant (start/stop) frames. This index will also contain information on the digital or video nature of each frame; i.e., video frames, when accessed, will be directed to a video monitor; the output from digital frames will be directed to the computer and then to whatever output device is desired.

1. Lister Hill Center Program Although Kenney [14] identified the potential for the storage and retrieval of digital information on optical videodiscs in 1976, there has been no evidence of private sector developments in this area. If realized, however, this application of the technology is of great potential importance to the IS&R community. It can serve as a publication process for large databases and full text document collections, including mixed video and digital information. Towards this end, the Lister Hill National Center for Biomedical Communications of the National Library of Medicine has initiated an R&D program to address the encoding of digital information on optical videodiscs and its playback via the industrial-type players. The theoretical, error free storage densities for this process are estimated to be between 20 and 30 billion bits or 2 to 3 billion characters per side. The usable data per side will depend upon the types and magnitude of the errors to be encountered in the mastering and replication of the digitally encoded videodiscs and the error correction methodologies to be employed. An experimental facility is being assembled at the Lister Hill Center to allow for the identification of such

errors. Investigations will seek to determine the effect of current videodisc production standards on information storage; viz., to what extent is the theoretical storage degraded by errors encountered in mastering and replication? Subsequent investigations will seek to identify where improvements in the production process will have the greatest impact on usable information storage.

2. Videodisc Interface Unit As part of this R&D program, the Lister Hill Center has developed an intelligent videodisc interface unit (VIU) to allow computer control of industrial-type videodisc players [15]. While numerous investigators are interfacing industrial-type players directly to microprocessors or computers, the VIU offers improved capabilities, a set of higher level commands for controlling the player, and a degree of device independence. The latter is achieved by a modular design such that different industrial-type players may be interfaced without changing the higher level control commands used by the computer program. In production, the VIU should cost between $600 and $700. The VIU is presently programmed to interface the DVA Model 7820 Industrial Player with either a terminal or host computer.

V. IMPLICATIONS FOR IS&R

The optical digital disc and the successful application of the optical videodisc to digital storage promises to impact online information retrieval services in a variety of ways, some complementary and some competitively. The optical digital disc will not only be an important adjunct to large centralized services, but also to distributed minicomputer based systems where purchase of the requisite disc drives can be justified. The videodisc, with its potential for publication of machine readable data bases may provide the greatest competition for centralized, online services. Even here, though, the local provision of graphics, audiovisuals, and full text may be complementary rather than competitive. The most conspicuous impact of the optical digital disk will be in terms of online storage costs. In addition, the very scale of online storage which could be realistically made available will, by itself, generate new offerings and services.

A. Random Access Graphics

One important application of the optical videodisc will be the local provision of graphics to complement online information. Online information delivery has traditionally suffered, in comparison with printed publications, by the absence of graphic material. The desirability of including graphic material extends from simple line drawings, such as chemical structures, to more complex line drawings, such as in patent submissions, to half-tones, color images, and full audio-visual sequences. The latter will be of particular importance as a complement to full-text encyclopedic data bases and computer-based educational material.

Pergamon International publicly demonstrated in March, 1981, an online system, Video PATSEARCH (15), for searching the 700,000 USA patents certified since 1971. Subscribers to PATSEARCH will obtain a special intelligent terminal with attached videodisc player (presently a DVA model 7820) and eight videodiscs containing the drawings of all 700,000 patents. The USA patent database is accessed by a user through an online vendor, BRS, Inc. After logging the terminal into the online database, and mounting the appropriate videodisc, the user may call up, alternately, the patent text (citation) or the corresponding patent drawings via the local videodisc. The Pergamon terminal also has sufficient intelligence to allow a tailored user-cordial interface [17] which minimizes the knowledge the end-user requires to effectively negotiate the necessary search strategies and access the corresponding drawings. The Canadian Patent Office also plans to install and test VIDEO PATSEARCH [16].

VIDEO PATSEARCH is the first commercially marketed application of videodisc-supported graphics for online IS&R. It is certain to be followed by many other applications such as catalogs of satellite imagery, museum collections, medical illustrations, etc. Each videodisc will have a capacity for approximately 100,000 images.

B. Online Digital Storage

1. Economics Economic savings may be realized by a reduction in the media cost, in equipment, and in physical floor space. While many discussions of computer memory technology emphasize the cost of the media, the cost of online storage is primarily a function of equipment (e.g., disc drives). In fact, some of the newer magnetic disc storage systems such as the IBM 3370 and the CDC 33502 are fixed disc systems and, hence, the medium is integral with the drive. Even the floor space, herein estimated at $25 per square foot, or $600 per disc drive, is small when compared to the cost of the drives, as shown in Table I. The cost for the IBM and CDC disc drives were obtained from USA Government Schedules and are to be used only for comparison. In each case, the cost of the control unit was averaged over the maximum number of drives that could be attached. The costs in Table I do not include the additional equipment in the central processor (e.g., extra I/O channels or special interface units) needed to support additional control units.

Table I, therefore, provides a gross picture of some economies that could be expected to accrue from the present and near future storage densities of the optical digital disc. Online storage can account for 50 percent or more of the total equipment costs used by an online service. The NLM maintains in excess of 10 billion characters of online storage. Even so, the number of online databases maintained by NLM is far fewer than those maintained by the major online services such as SDC and Lockheed. As shown in Table I, the

TABLE I. Cost Comparisons for Online Storage

		Mbytes	$K/Drive	$K per 10 Gbytes*	$K per 100 Gbytes
1973	IBM 3330-11	200	28	1,430	14,300
1975	IBM 3350	317	43	1,300	13,000
1978	CDC 33502	635	27	430	4,300
1979	IBM 3370	571	29	480	4,800
1981	IBM 3380	2,500	98	392	3,920
1982**	Philips Disk	2,000	20**	100	1,000
1983**	Philips Pack	100,000	50**	51	51

* gigabytes or billion characters
** estimates

most dramatic economies to be foreseen are still over a year away and are predicated on the availability of a 1 trillion bit (100 billion character) optical disc pack.

2. Access to Full Text One qualitative trend, given the availability of low-cost mass storage systems such as the optical discs, will be the greater online availability of full text. Estimates for the transition of scientific journals and documents to full electronic printing (including graphics) are 10-15 years. However, the technology presently exists for storing journal copy in high resolution, compressed facsimile format to support online browsing or computer-driven demand publication. Another major program effort within the Lister Hill Center, "The Electronic Document Storage and Retrieval Program", is directed towards the intgeration and further development of these technologies. The high storage density and random access characteristics of the optical disc are driving technology developments in this direction.

Where the full text is available in machine readable form, the optical disc will not only provide an economical storage medium, but has the additional potential for filling a long standing gap in the effective utilization of associative text processing. Previous efforts in associative text processing have concentrated on the computer processors. A major limitation has been the lack of relatively low-cost, random access storage units with both sufficient capacity and data rates. Optical digital discs, and disc packs, combine both the high data rates and large storage capacities required for these applications. Recent developments [18] have demonstrated data rates of 30 Megabits per second with capacities in excess of 30 billion bits per 12 inch disc, and data systems with data rates of 60 Megabits are shown to be feasible.

Just how associative text processing will best be employed - for online access to text, or preprocessing of text for the creation of indices, or both - is still unclear. It is only certain that the availability of storage technologies such as these will give impetus to their reconsideration.

Another potential application will be the local or regional provision of full text (with graphics) as a complement to online, centralized searching. The latter implies high speed, sophisticated processing requirements which may be best centralized, while the provision of audiovisual segments and full text with graphics to the remote user implies major communications requirements and may best be decentralized. Since communications are, in most countries, tariffed, the economic trade-offs between searching and information delivery may not be dictated by technology alone.

C. Distributed Vs. Centralized Information Delivery

The potential for distribution of machine readable information via optical videodiscs or optical digital discs may appear as a threat to centralized, online services. Given the alternatives, decisions will be based on considerations similar to those dictating the choice of in-house vs. time-shared computing services. These considerations are primarily economic and include the cost of computer equipment, communications, personnel, and the frequency of need for the information at the going market rate.

For the distributed delivery of full text information, the difference between the optical digital disc and the proposed digital application of the optical videodisc includes both publication and equipment cost. The optical digital disc can, at present, be replicated only by the sequential copying from one disc (or pack) to another. This should pose no major impediment to regional distribution of information, but would be very costly in quantities of 100 or more copies. Mass publication in such quantities (or larger) would be in the domain of the optical videodisc technology. Similarly, for single users, exploitation of the commercially available industrial players may greatly reduce equipment costs. Such players, on the other hand, will not effectively support multiple users.

1. Microform Surrogate As soon as the storage of journal and document pages is considered, we are, by implication, addressing a potential surrogate for microforms. While the optical discs will have better random access capabilities, will they compare economically with microforms? In order to obtain such a comparison it will be necessary to make some assumptions regarding the amount of digital storage required per journal page. Although normal facsimile reproduction employs a scanning resolution up to 200 lines per inch (40,000 bits per sq. in.), this is insufficient for high fidelity archival storage of journal pages. It is estimated that for good reproduction of text (not half-tones or color graphics), approximately 500 lines per inch (250,000 bits per sq.in.) will be required to accurately reproduce, with excellent quality, the smallest characters found in scientific journals. Estimates for digital compression for these resolutions range from 20:1 to 100:1. Hence, assuming a scanning resolution of 500 lines per inch, a 50:1 compression ratio, a page size of 8" x 10", and a disk storage of 20 billion bits, will result in 50,000 stored pages per disk. Depending on quantity, the videodiscs will cost between $2 and $10. Hence, cost will vary between $.04 and $.20 per 1,000 pages. By comparison, a standard 24x microfiche contains approximately 100 frames per transparency, while the NCR ultrafiche, as used by the British Library for their "Books in English" publication [20], contains 2,380 frames per transparency. Assuming a cost of fiche duplication at $.25 per transparency for standard fiche and $.50 per transparency for ultrafiche, this results in costs of from $.20 to $2.50 per 1,000 pages. Hence, in terms of the media, the videodisc may be an economic competitor to microform for large collections.

It must be noted that the difference in equipment costs will preclude competition between microforms and optical discs for the individual user for the foreseeable future. A microfiche reader can be very inexpensive. Access to information stored on optical disc will, on the other hand, require optical disk players, high resolution displays and/or printers, and some form of computer (or microprocessor) capability. It can be expected, nonetheless, that videodiscs will compete with microform in (a) regional centers that subscribe to major microform document collections such as ERIC (Educational Research Information Center) and provide for local document delivery by microform duplication and hardcopy blow-back, and in (b) libraries and information centers that have cost-justified the acquisition of the appropriate equipment. It will also be possible to generate microforms by COM (computer output microform) directly from the optical disk storage.

2. Document Delivery In a previous section, the optical digital disc was discussed as a potential document storage medium. There exists a natural extension to this in terms of document delivery. Given the facsimile storage of the documents and journals as described, there exists today computer-driven xerographic laser printers by IBM, Siemens, and Xerox that could print this information at rates of up to 2 pages per second (ca. 14 million pages annually per machine per single shift of 40 hours per week). Of these machines, the Xerox 9700 laser printer has, at present, the highest resolution of 300 lines per inch. A version of this machine, modified to handle card stock, is employed by the Library of Congress to publish catalog cards from the MARC database. While these machines are expensive (ca. $350,000), high volume output brings the cost per page to 1 cent or less. It is safe to say that, in the future, a spectrum of lower cost, albeit slower, printers will be available for distributed printing.

Towards this end, whether the documents are printed centrally, or the information is transmitted over document delivery networks to local printers, or distributed via videodiscs for local access and printing, will be a matter for future determination. The deciding factors may be political/economic rather than technical; e.g., the establishment of communications tariffs.

Note, however, if centralized distribution is utilized, then videodisc publication is not required and the higher storage densities of the optical digital disc could be employed. Extending our previous calculation for the page storage to the higher densities, and assuming 1,000 pages per journal per year, the figures in Table II are obtained.

TABLE II. Estimated Page Storage Densities

		Billion Characters	Pages	Journal Volumes
1980	Disk	2	50,000	50
1982*	Disk	20	500,000	500
1983*	Disc Pack	100	2,500,000	2,500

*Estimated

3. Hybrid Network Architecture In order to clarify the concept of a hybrid network architecture, as intended here, the example of library shared cataloging will be explored:

It is not reasonable to take pictures of library catalog cards for storage as video frames on a videodisc because of the lack of resolution on standard video monitors. If one assumes, however, that a catalog entry has, on the average, 500 or less characters, then a video or optical disk, with a storage capacity of 2 billion characters could store

the contents of 4 million catalog entries.

While there are very few individual libraries with 4 million catalog entries, all libraries may have the need to share access to machine readable, regional and national union catalogs. This potential of inexpensively duplicating such large collections of catalog information may result in new approaches to shared cataloging. For example, the consideration of network strategies in such areas as shared cataloging may be seen to result from two conflicting premises:

1. To maintain currency, the catalog should be maintained and updated centrally.

2. It is impractical and undesirable to service a nation of online users from one, centralized computer center.

Regarding the first premise, it is obviously desirable that a librarian wishing to catalog a new acquisition have access to the efforts of another librarian who has cataloged that item even minutes or seconds before. Regional network strategies have been recommended to resolve the conflict between these two premises. In such strategies, the database would be replicated at each regional center to service online users in each region. The regional centers would maintain currency among themselves.

The potential of utilizing the optical videodisc as an inexpensive publications medium could, in the future, provide an alternative network strategy. This alternative strategy may incorporate three hierarchical levels: the central update facility, a local (or regional) storage of "recent" additions, and a cumulative, historical collection on videodisc. In such a strategy, the optical or videodisc-based historical collection will be published periodically (monthly, bimonthly, quarterly, etc.). The "recent" storage will be standard read/write storage and will be updated via the network, weekly, nightly, etc. from the central facility. A catalog query to the site serving the user will first search the historical and "recent" catalogs. If no local information is found, then the request will be automatically routed to the central facility. The frequency of videodisc publication and "recent" catalog updates will be chosen to maintain a reasonable rate of transactions at the central facility.

Hence, hybrid network architectures may include access to distributed historical and recent data, while, at the same time, provide centralized access to the most current data.

D. Office and Local Archive Systems

1. Systems The DRAW system, pioneered by North American Philips and discussed in Section III.A utilizes relatively powerful gas lasers and sophisticated tracking or indexing mechanisms to insure precise positioning of each track. Two other systems are under development that utilize "pre-grooved" optical discs. These systems allow less expensive and more robust tracking mechanisms and are designed for local office environments. One system has been under development by Philips-Eindhoven [19] for a number of years. The other, by Toshiba (Japan) [13], was demonstrated at the 1981 National Computer Conference and should reach the market within two years.

2. Uses Both systems are intended to support the archival storage of local correspondence and other office documentation. The Toshiba unit contains integrated facsimile and laser printing unics. The former to scan document images which are then stored on disc, and the latter to print documents on demand from a disc. The resolution presently employed is approximately 200 lines per inch (7 lines per mm.). While sufficient for typewritten material, it is not sufficient resolution for archival quality journal page images, hence the emphasis on office systems. Such systems should, however, provide good archival storage and source data capture for many of the applications discussed in the Artemis report [20].

E. Preservation

A major advantage of optical discs is their potential archival quality. Accelerated aging in the laboratory has been inconclusive in determining actual lifetimes. At present, lifetime appears, for all practical purposes to be unlimited for videodiscs. Oxidation of the substrate in Philips-type DRAW discs was discussed in Section IV.A.2, as well as new developments that may eliminate this potential problem such as the DREXON draw discs.

Even if the discs do not provide unlimited archival lifetime, the manner in which the data is stored (digitally) will allow periodic regeneration of the discs without the attendant loss in quality now associated with different generations of images on photographic film. This "regeneration without loss of quality" will be an expected occurrence as improved materials are found.

VI. Conclusions

Optical disc technology portends many dramatic improvements for IS&R in the Eighties. It should also be clear from the preceding discussions that the promise of the optical disc technology for digital storage is the driving force for many complementary technologies such as high resolution digital scanning and computer-driven laser printing.

The optical disc, itself, is an almost ideal medium for IS&R. If it were necessary to choose between read/write or read/only-archival storage, all other aspects being equal, the latter may be far more important. User access to online information is, for the overwhelming majority of cases, read-only anyway; i.e., the user is neither allowed nor does he have a need to change the source of the information.

REFERENCES

[1] C.M. Goldstein, "The Potential Impact of Optical Disk Technology on Libraries and Online Services" in "Telecommunications and Libraries; a Primer for Librarians and Information Managers" Knowledge Industry Publications, Inc., White Plains, NY; To be published in June 1981

[2] R.A. Bartolini, A.E. Bell, R.E. Flory, M. Lurie, and F.W. Spong, "Optical Disk Systems Emerge" IEEE Spectrum, pp. 20-28, August 1978.

[3] C. Bricot, J.C. Lehureau, and C. Puech, "Optical Readout of Videodisc" IEEE Transactions on Consumer Electronics, November 1976, pp. 304-308

[4] J. Winslow, "Mastering and Replication of Optical Videodisc" Journal of Optical Society of America, p. 379, April 1976.

[5] G. Kenney, D. Lou, P. Janssen, J. Wagner, F. Zernike, R. McFarlane, and A. Chan, "An Optical Disc Data Recorder" Philips Laboratories Report, Briarcliff Manor, New York. 13 pp.

[6] G.C. Kenney, D.Y.K. Lou, R. McFarlane, A.Y. Chan, J.S. Nadan, T.R. Kohler, J.G. Wagner, and F. Zernike, "An Optical Disk Replaces 25 Mag Tapes" IEEE Spectrum, pp. 33-38, February 1979.

[7] J. Drexler, "Applications of a New Laser Recorded Media for Digital Data Storage", Proceedings of the Electro-Optics/Laser 79 Conference, Anaheim, California, pp. 69-78, October 23-25, 1979.

[8] J. Drexler, "DREXON [TM] Disc Storage for Laser Printers" Society of Charter Optical Instrumentation Engineers Proceedings, Vol. 169, pp. 11-16, Laser Printing, 1979.

[9] J. Drexler, "DREXON optical memory media for laser recording and archival data storage" Proceedings of the Symposium on Optical Storage Materials, November 1980, Ossing, New York

[10] J. McLeod, "Metal-plastic prolongs optical-disk lifespan" Electronic Design, August 2, 1980

[11] "Semiconductor Laser Simplifies Optical Videodisc Player" Electronics, May 26, 1977.

[12] M. Marshall, "Optical disks excite industry" Electronics, May 5,1981, pp. 97-102

[13] W. Thomas, "Video Disk System Comes Out of Philips Laboratory," ComputerWorld, 1 p., Jan. 23, 1979.

[14] G.C. Kenney, "Special Purpose Applications of the Optical Videodisc System" in Philips Laboratories Report, Briarcliff Manor, New York. Also in IEEE Transactions on Consumer Electronics, pp. 327-338, November 1976.

[15] W.D. Russell, W.H. Ford and C.M. Goldstein, "A Videodisc Interface Unit" Lister Hill Center Technical Report 27 pp., January 1981.

[16] "Pergamon International Introduces Patent Search System" R&D Management Digest, L.H. Hattery, editor, V.10, No.10, pp. 1-2 April 1981

[17] Goldstein, Charles M. and Ford, William H. Jr., "The User-Cordial Interface," Online Review, Vol. 2, No. 3, 1978.

[18] R.A. Bartolini, A.E. Bell, F.W. Spong, "Diode laser optical recording using trilayer structures" IEEE Journal of Quantum Electronics, V. QE-17, January 1981, pp. 69-77

[19] K. Bulthuis, M.G. Carasso, J.P.J. Heemskerk, P.J. Kivits, W.J. Kleuters, and P. Zalm, "Ten Billion Bits On A Disk" IEEE Spectrum, pp. 26-33, August 1979.

[20] "Artemis: A System for Document Digitalization and Teletransmission", Report to the Directorate General XIII of the Commission of European Communities, April, 1980, Arthur D. Little Rept. 83405

USERS REQUIREMENTS FOR TERMINALS; SELECTION, USE AND ERGONOMIC FACTS

H.J.LEEBEEK,
Institute for Perception TNO
Soesterberg,
THE NETHERLANDS.

INTRODUCTION

In many process-control situations a computer is used to assist the man operating the system. The various tasks can then be allocated between man and machine: the machine, or the computer, to undertake the dull, repetitious or fatiguing tasks; the human operator to undertake the supervisory and decision-making tasks.

The elements of the tasks thus designed and assigned to the human operator are the perception of the machine output signals and the initiation of a desired state through control actions. Such man-machine interfaces can be described by the cycle given in Fig. 1.

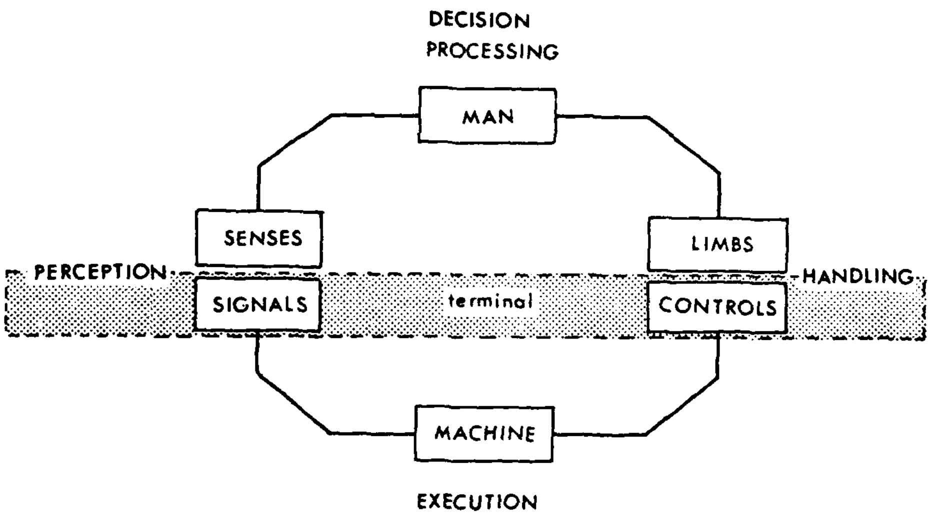


Fig. 1 The ergonomic man - machine cycle

The machine presents information to the operator, who must perceive it; then process the information and make a decision. The operator must implement the decision by control actions using his limbs. The machine must then respond to the control action and the new state shall be indicated by signals which again must be perceived by the operator.......etc.

This chain is as strong as its weakest link. No part in the cycle can be neglected. Nevertheless man machine cooperation often fails by bad adaptation of the machine to human capabilities and limitations. This occurs in the man machine interface: the terminal.

Ergononomists study these adatation aspects and they can give a great many recommendations and rules of thumb for an optimal cooperation between man and machine.

In Fig. 1 a horizontal zone in the man-machine cycle is drawn. It represents the function of a terminal in this cycle. It is clear that terminals have output and input devices. Some of these possibilities are shown in Fig. 2.

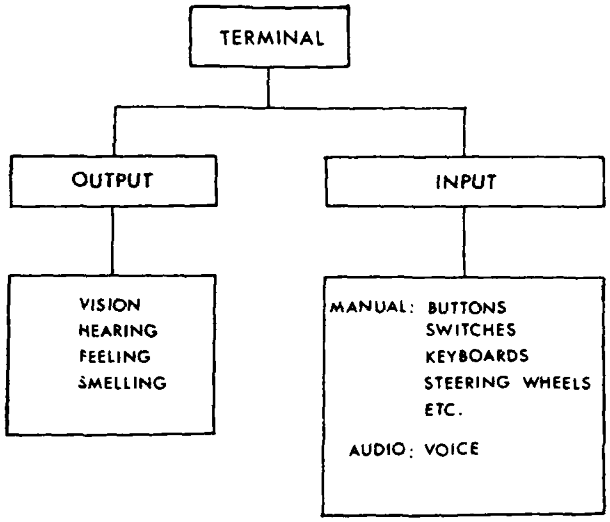


Fig. 2 Typical output and input devices of a terminal

All input and output devices ought to be adapted to the operator of the terminal. It would be a shortcoming of the system if the operator had to be adapted to the terminal.
In this paper typical human aspects of terminals will be discussed.

In many cases conventional processing and machine handling is replaced by computer handling. This paper will discuss the typical new computer terminals. At first a description of the present state of engineering will be given, also an expectation of future equipment will be briefly mentioned.

There is a tendency to use more keyboard input devices instead of a great diversity of buttons, switches, wheels, etc.
So it is unavoidable to study workstations with visual display units connected with keyboards.
Very often people think that these devices are so new and so atypical that recommendations for conventional equipment fail to cover the problems.
However most conventional recommendations and rules of thumb can be directly applied to modern devices like VDU's.

KINDS OF TERMINALS

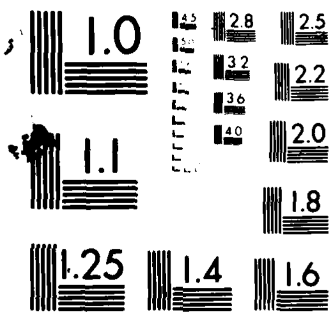
Several types of terminals can be distinguished with respect to information presentation. Fig. 3 gives further details of the possibilities of typical computer terminals as well as more conventional devices.
A division can be made between 1. auditive information output: bells, buzzers, sirens, etc. and
2. visual information output: a) signal lights, instruments, dials, etc.
b) symbols, graphs, etc.

At the end of the paper some remarks on auditive input and output will be given. However the main part will be a discussion on typical visual aspects.

AD-A111 846 ADVISORY GROUP FOR AEROSPACE RESEARCH AND DEVELOPMENT--ETC F/G 5/2
WHAT SHOULD USERS EXPECT FROM INFORMATION STORAGE AND RETRIEVAL--ETC(U)
DEC 81
UNCLASSIFIED AGARD-CP-304 NL

2 of 2

END
DATE
FILMED
4 82
DTIC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS 1963 A

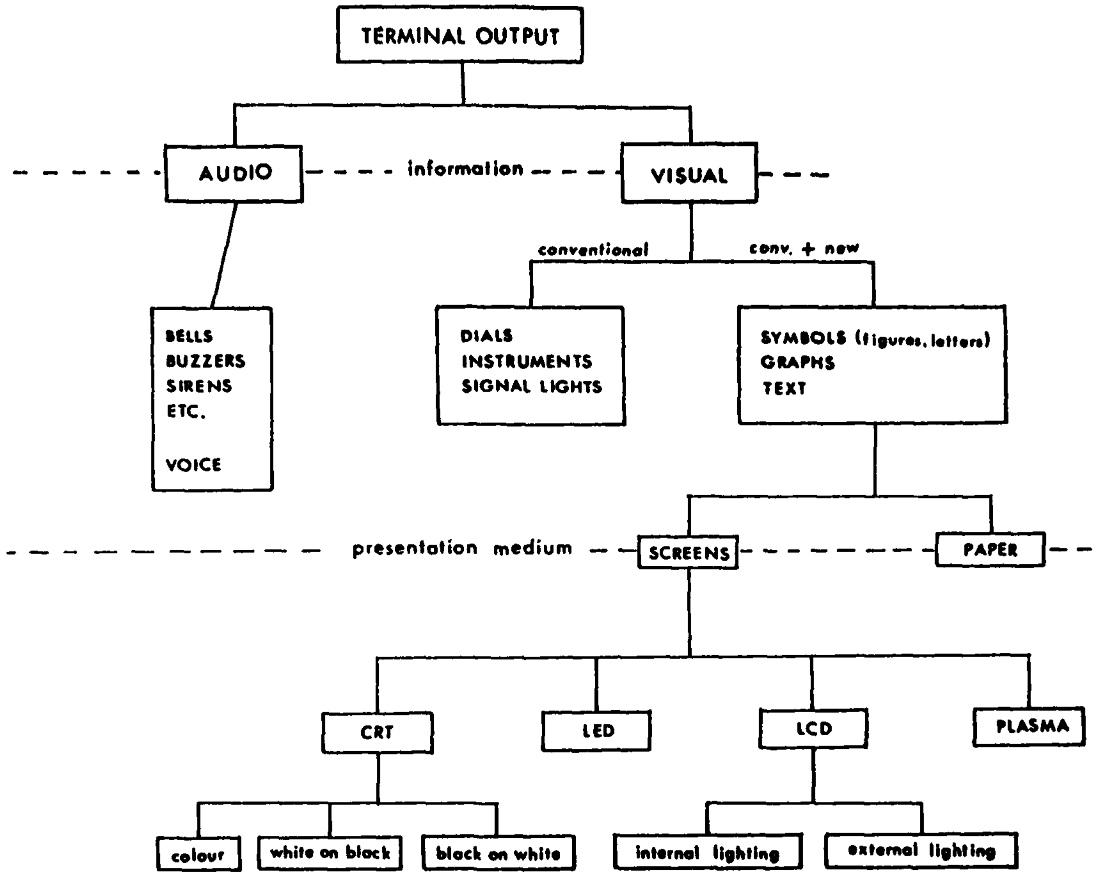


Fig. 3 Kinds of terminal output

CRT = Cathode Ray Tube

LED = Light Emitting Diode

LCD = Liquid Cristal Display

VISUAL ASPECTS

It is possible to base recommendations and rules of thumb for visual information presentation on a description of the visual field capabilities of the human observer.

An image of the environment is present on the eye's retina. The characteristics of the eye, including the retina and the nervous system behind the retina, define detection and recognition of objects in the visual field.

<u>Contrasts and luminances</u>

The perception of objects depend on their luminances, but within certain limits the contrast (= luminance ratio) between the different parts of the field of view defines the probability of perception.

A rule of thumb can be given:

Table 1 Contrasts

Luminance ratio $\frac{L_{high}}{L_{low}}$	Subjective contrast impression
1	no contrast
1 - 3	low contrast
3 - 10	medium, well visible, contrast
>10	(too) high contrast

Normal contrasts are all in the range 3 - 10. Examples: In a newspaper between white paper and black printing ink, 7; high quality printing on high quality white paper, 10; lead pencil on white paper, 3; symbols on average CRT, 3 - 100: that often means too high.

The luminance of self-luminous surfaces is almost independ of the illumination of the work station. In these conditions high contrasts may exist.
The luminance of externally lighted surfaces depends on the illumination of the work station; the contrasts are mostly less high than for self-luminous surfaces.

Specular reflections

In a window pane or in a glossy surface specular reflections can cause glare. The image of the lighting fitting or other objects with high luminances is so strongly visible that it causes a certain annoyance with respect to the proper object behind the pane or under the glossy layer.

When the contrast between proper object and the annoying image is greater than 5 - 10 the annoying glare will be only slight or negligible.
Some information presentation media are so prone to specular glare that one would better not use them.

Visual acuity and Size of detail

The human eye is able to see an object having an angular size of 1 minute of arc. However rather many people have lower visual acuity because of not using their glasses, refusing to wear glasses or not knowing that they can get a better visual acuity.
So for practical use it is better to base the information presentation on a minimum critical detail of 2 minutes of arc. This means a detail of 1 mm at a viewing distance of about 1.5 m.
The height of symbols (letters and figures) must have at least 5 critical details and in general the width must have 3 critical details. Thus the ratio $\frac{\text{width}}{\text{height}}$ shall be at least 0.6. In oblique viewing a greater or smaller ratio is recommended dependent on the direction of viewing.

The legibility of a symbol increases when it is built with more details and when the details itself are not too large. Some presentation media, like CRT's etc., build their symbols with such large details that the legibility is less than the optimal.
A normally printed symbol can be seen at a viewing distance of 300x its height. For a 5 x 7 point matrix symbol a viewing distance of 150x its height is recommended.

Review of properties and suitability

Table II gives a review of the different information devices.

Table II

	CONVENTIONAL		CONVENTIONAL + NEW							
	DIALS	SIGNAL LIGHTS	CRT			LED	LCD		PLASMA	PAPER
			colour	white on black	black on white		internal light	external light		
contrast in task 1)	O	X O -	X O	X O	O -	X	X O	O -	X	O -
contrast task - surrounding 1)	O -	X O -	O -	X	O -	O -	X O -	-	X	-
specular reflections annoyance 2)	X	O	O	X O	-	O	-	X	X	X O -
minimum size of detail 3)	X O -	X O	X	X	X O	X	X O	X O	X	X O -

X means: 1) high 2) much 3) large

O means: 1) medium 2) medium 3) medium

- means: 1) low 2) little 3) small

Accommodation range of older people

In Fig. 4 the ranges of accommodation as they are available for people of different ages are drawn.

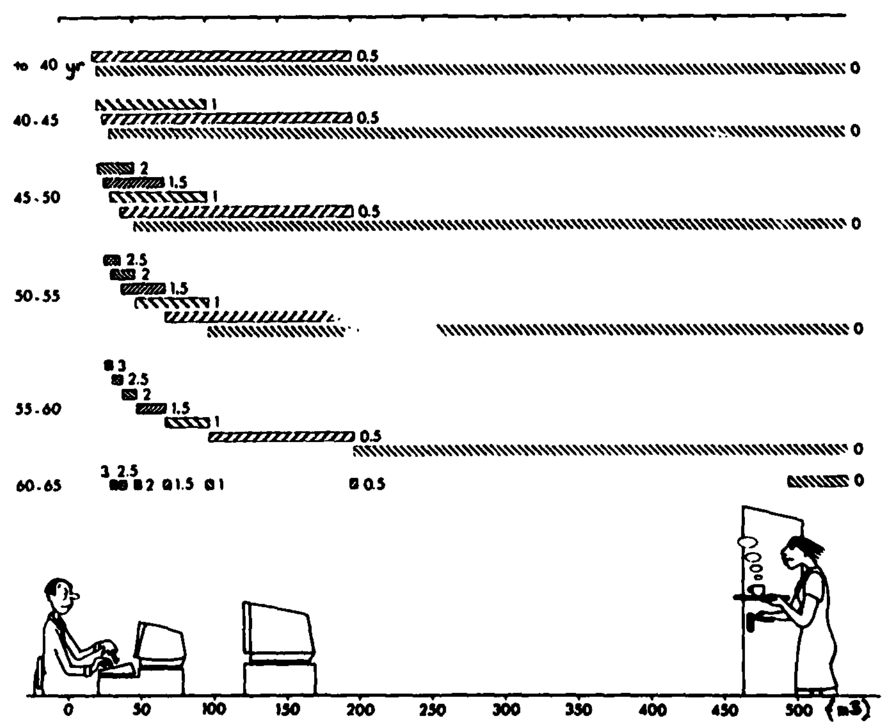


Fig. 4

Horizontally, the viewing distances are given; along the vertical axis, the different age groups. For each age group horizontal shaded bands indicate on which viewing distances focussing by means of glasses is possible. The glasses diopters are written next to the bands.
Higher ages require higher diopter glasses correction to focus at short distances and this means a decrease of the viewing distance range on which it is possible to focus with just one diopter correction.

Warning: This figure is no glasses prescription. The data given are mean values and cannot be applied for individual persons.

OTHER TERMINAL ASPECTS

Working posture

In many work stations bad working postures are encouraged.

Recommendations: Keyboards shall be as thin as possible (max. thickness 2.5 cm).
Keyboards shall be placed as low as possible.
Screens shall be placed on a height corresponding to a line of sight of 15 degrees below the horizon.

Audio aspects

Audio information output is easier to perceive than visual output. However it attracks the attention of the operator in such an intrusive way that oudio output is mostly restricted to alarm situations. Future developments give reason to think that voice input and voice output may have a greater application in future automation processes.

The Use of Small Computers in Information Retrieval from Textual and Free-Structured Data-Bases.

J. H. Ashford, BSc, PhD, FGS, MIInfSc.
J. H. Ashford Associates Limited
72, Harrow Lane, Maidenhead, SL6 7PA

Summary Minicomputer systems are well established for the storage and retrieval of bibliographic records and the full text of documents. Examples are given of a range of applications including the handling of numeric and experimental data. The relationship between free text and structured records is examined, and the appearance of hybrid data forms is noted.

The potential for the use of microprocessor based systems is not yet well developed, and is limited by the relatively high cost of storage devices, and by the difficulty of providing flexible and reliable software. The most promising developments are following the minicomputer systems in the use of 'package' software, and in having close links to word processing systems.

Future developments are likely to centre on reducing costs of storage; substituting associative processors for some of the functions now performed by software; development of user interfaces, and new methods of search planning and optimisation. The use of communications links, both locally and over long distances will be of increasing importance as 'electronic publishing' becomes accepted for a wide range of information.

The position in 1981 A survey was made in 1978 - 79 of the potential market for package free text storage and retrieval systems in UK; this was published in summary (1) and later more fully (2). One of the main conclusions of the study was that '25% of respondents expected to be using some form of free text information system by end 1981'. The main application areas identified were bibliographic records; full text of technical documents; commercial information; and special applications such as legal case studies, and accident and incident databases. With the benefit of hindsight, it can now be seen that while the survey correctly identified the high rate of new interest in text handling systems, it very much underestimated the importance of word processing developments and the significance of equipment acquired for document preparation as a minimal cost route to database acquisition for local systems. When the studies of potential markets for packages were extended to North America, a new and important application was identified - the build up and exploitation of document databases in support of litigation in major civil proceedings. The main application areas of free text systems for information storage and retrieval on small computers are now believed to be:

- Bibliographic records
  - Catalogues of special libraries *
  - Catalogues of report and document collections *
  - Bibliographic records and abstracts for abstracts journals etc. and related information services *
- Full text systems
  - Legal statute, cases and related documents
  - Documents in support of litigation *
  - Technical reports, patents etc. *
  - Laboratory and clinical records in pharmaceutical development and a number of similar 'health and safety' applications
- Special systems
  - Special library circulation control systems *
  - Hazardous chemicals databases
  - Commercial records systems - markets, prospects, proposals, invoices, servicing reports etc. *

The applications marked * above are particularly developed on minicomputers; most of the others, with the exception of the very large shared legal databases, are potential small machine systems. A selection of operational systems is listed in Appendix 'A' to illustrate the range of applications, hardware and software which have been successfully implemented. The online information vendors, offering 'dial-in' access to large bibliographic databases have not in general used minicomputers; there is, however at least one 'financial information' service, the TEXTLINE system operating in the City of London, which runs on a DEC PDP 11/70 minicomputer.

At the time of writing - June 1981 - very few microcomputer based storage and retrieval systems for text information have progressed beyond small scale or experimental applications. Those of which the writer is aware are discussed later under current developments.

A useful distinction can be made between the textual and free structured databases, with which this account is mainly concerned, and the formally structured database systems used largely in commercial and 'numeric' data processing, based on the way in which stored records are indexed and accessed by the user. The usual approach taken with free text systems is to form an inverted file of all (or most) of the 'words' found in the text. This inverted file, or concordance, then points back to the source articles from which each word was drawn, often at a level of 'chapter; article; paragraph; word', and is used to support the processing of boolean query statements containing one or more words as arguments. This method allows reference to the database to be made on the content of the articles, treating all words as tokens of equal status. The structured database handler, on the other hand, uses a predetermined format to allow definition of the role of a field, such as 'Department' or 'Cost Centre', so that processes such as listing all Departments in sequence of Department Number are naturally facilitated. (This subject is developed at length in (3) and (4)).

The simplest form of free text record is represented by a wholly unstructured 'article' of continuous text. This is rarely used, but the next stage, the Title plus Abstract form is very common. Here only the Title field is role defined; all other matter, including author, source etc. is treated as continuous text with the abstract. In a bibliographic record - as used in a catalogue in a special library, for example - it may be desirable to distinguish words used in the separate contexts of Title; Authors; Classmark; Location; Date; Volume; Publisher; Pagination and format; Current borrower; Reservers. These role subdivisions of free text records are achieved by setting up a loose structure of named sections, where the section is used as a limiting 'window' to narrow the access to part only of the data. Few free text applications use records so much subdivided as the MARC International Exchange Format for catalogue data; however it is becoming usual to find five or six sections in bibliographic records; five or more in commercial record systems; up to ten in personnel records. Thus the free text database handler has moved towards the role oriented approach of the structured database, although without gaining the flexibility of amendment associated with preformatted 'pigeon-hole' like structures.

A recent and interesting response to the need for partial structure combined with flexibility in amendment of rapidly changing data fields has been the development of hybrid databases, where part of the data is held on a free text file, and accessed by content, and part is held as a series of formatted records in one-to-one correspondence on a conventional indexed sequential file. This combination is seen when, for example, circulation control, with its many small transactions, is added to the handling of a library catalogue on a free text system. The link between the two parts of the hybrid is sometimes made directly within the software and is not under control of the user, sometimes through a visible connecting reference number contained in both parts of the record and treated as a user option. Other hybrid systems of this sort are beginning to be used in handling of membership record of societies, and in the coupling of maintenance records to textual databases describing plant and equipment.

For textual databases of medium scale, say ten million to one thousand million characters of stored information, the use of minicomputer systems is well established. Many of these systems use software package products, alone or with local adaptations, as the nucleus of the operation. With the exception of academic institutions, where software is often written from first steps which could readily have been had from outside vendors, the package approach to database software is gaining increasing acceptance, not only for the cost savings it offers, but for the assurance of reliability and adequate documentation and maintenance. In the course of discussion at a recent conference (5), over twenty free text retrieval systems were identified which had been used successfully on a number of sites. About one third were available on minicomputers, but the proportion of recently developed systems using minicomputers was much higher.

Small databases, on the other hand, are being treated in all sorts of ways. Some have been fitted into relational database handlers (such as INFO, RAPPORT); others are kept on mainframe systems using free text software; still others are managed on bureau services. However, one of the most rapid recent developments has been the adaptation of word processors to act as storage and retrieval systems of varying sizes and degrees of sophisti cation. A detailed account of a carefully thought out series of applications describes word processors in an industrial environment (Pilkington Glass Limited) to provide both local simple database systems, and a facility for information transfer between a number of widely scattered sites. (6).

Current developments Three main areas of development activity are running in parallel; the tidying up and consolidation of the 'main stream' free text package systems, both on mainframes and minicomputers; the development of preprocessing and postprocessing utilities to make the storage and retrieval systems easier to use; and an epidemic of microprocessor developments of all sorts, many of very doubtful value, but some likely to be the stem for an important range of low cost, local or personal database processors. The availability of cheap microprocessor hardware, and relatively cheap 'floppy' disc storage up to perhaps two million characters online at a time is the main current stimulant from the equipment standpoint. Software development is concerned more with marginal improvements and utility programs than with radical innovations, at least on a scale of two years ahead.

The consolidation of existing technology is well illustrated by two of the package systems available, or to be available on minicomputers. The first, STATUS, has been in use since 1970, and in its considerably evolved present form since 1980. The current version, STATUS 80 is regarded by the development group as having reached a reasonable limit to its range of features, and work on the central system - as opposed to its

ancillary and auxiliary programs - is concentrating on further gains in run time efficiency, better use of storage space, and adaptation to more machines and wider markets. SIFT, on the other hand, is a new (7) package system being developed by the Statens rasjonaliseringsdirektorat in Oslo, Norway, to provide a software vehicle for legal databases in the first instance. SIFT is a fairly advanced derivative of an early version of STATUS, but has undergone a good deal of local revision and development. The point of interest for present purposes is that in this new initiative, there are no significant features not already available in STATUS 80 and its auxiliary software. (Nor, incidentally, is this true only of STATUS; there are several package systems available with essentially the same functional coverage - ASSASSIN, CAIRS, STAIRS, UNIDAS, for example available in Europe, and BASIS and STAIRS in North America). Most of the systems under active development are becoming easier to use and better organised - but no major new features are appearing. The technology is, therefore, now stable so far as the central software is concerned, and will probably not change much until new hardware developments change the operational context.

There is, however, a great deal of work being put in to software auxiliary systems to link the user to the system in such a way as to make it easy to use, or robust to errors in data, or generally more presentable. Most package systems, and many locally developed systems nowadays are aiming to provide:

Input format aid, allowing 'form-filling' input of text records;
Input of word processor texts;
Input from computer typesetting tapes; (not so common)
Output in tabular format;
Output to word processor;
Output to typesetter;
Automatic, or semi-automatic index generation;
Controlled language for keyword selection;
Thesaurus handler and links into searching.

In addition one might reasonable expect to find, either available or under development, routines for the efficient <u>bulk</u> take on of records from external sources; query reorganisation (unseen to the user) to minimise search times and comprehensive security and password control. Some systems, such as SIFT and STATUS now incorporate macroprocessor facilities, and in the latter, the macroprocessor permits the handling of conditional processes and run time specialisation, so that completely new 'verbs' can be added to the command language for particular users.

Now that a microcomputer with minimal storage on floppy discs can be bought for less than £500. and the output fed to a domestic television screen, there is no obstacle of any significance to the do-it-yourself programmer <u>except</u> the intrinsic difficulty of writing accurate, reliable, maintainable and functionally well conceived software. Since the difficulties are not obvious to the beginner, and since the satisfaction and excitement of getting one's first jobs to run are considerable, there are many thousands of man-hours now being spent on learning to program, and on producing amateur software of poor quality and limited utility. The minimum of tasks to be undertaken in any program development includes:

Definition of functional requirement;
Simulation check (desk check) that the functional specification is reasonable;
Design of data structure, taking account of storage efficiency, access efficiency, and, in the case of floppy discs, minimising wear during read or write cycles;
Design of program structure;
Writing of program code - preferably in a language with inherent structural features such as PASCAL or ALGOL or a structured version of FORTRAN;
Testing with synthetic data;
Testing with live data;
Exhaustive testing of error and exception cases;
Provision for security backup of the data;
Documentation of the system -
  for the user;
  for the programmer who will one day have to maintain it!

The writer is concerned that a major proportion of the microcomputer information systems now being written and proudly demonstrated will not survive the promotion or transfer of their authors. All this is not to say that one should not learn to program - but that software intended for regular use in situations where its performance matters, requires professional attention.

There <u>are</u> sound and well conceived microprocessor developments. The example cited above of Pilkington's use of word processors merits study as a good case of careful requirement specification carried through into practice with skill and judgement (6). Lundeen, in a recent paper, (9), cites a number of systems installed in libraries for cataloguing, circulation control and other applications which have been set up by software houses of good reputation. At least two fairly comprehensive free text storage and retrieval packages running on microprocessors will soon be available from teams with established reputations - one from Cuadra Associates of Santa Monica, Ca., and the other from the ASSASSIN team at ICI Agricultural Division in Billingham, UK. The best policy for a little while is probably to 'wait and see' and to greet innovations on microprocessors with reasonable scepticism for the time being.

<u>The future with small computers</u> The best method of forecasting developments in the computer industry appears to be to keep a list of those features one most wants which are just at the edge of credibility for seven years hence - and confidently expect them to turn up within three years if enough other people want them too. (This does <u>not</u> work for computer translation of natural languages; the travelling salesman problem; or a high level language so simple that the end user can specify his own applications without the need for DP types! These are special cases.) The following assertions must be taken to represent a mix of intelligent guesswork and 'wish list' for this writer.

The really useful hardware developments will be:

Cheap mass storage, possibly solid state, possible ever cheaper discs under competition from solid state devices. This will allow us to store very large databases at reasonable costs, for instance in the case of the full text of the proceedings of legislative assemblies - which are at present desirable but very costly. It will also, at the other end of the scale, make individual or local systems more viable economically.

Large scale, read-only data stores such as video discs will alter the economics of publishing, possibly making it more attractive to buy a database than to dial an online data vendor. Text and graphics will then become easier to interleave and electronic publishing might start to approach the aesthetic quality of traditional methods.

Hardware modules will replace heavily used software, including the embedding of parts of the retrieval programs in 'microcode' or equivalent; the use of associative stores to take over list or index searching tasks; and communications controllers such as X25 handlers embedded in hardware logic.

All of the above developments require fairly large numbers of users to want to do much the same sort of thing with their equipment. This is a normal feature of hardware innovation, and the buildup of a critical mass of users to make an innovation <u>economic</u> sometimes makes it look as though new ideas come in more of a rush than is actually the case.

Software innovations are much more of a problem to anticipate, partly because software does not have the investment lead time involved in mass producing cheap hardware, and partly because at any one time, a large number of researchers are pursuing parallel paths in different contexts, and it is the coincidence of a new theoretical construct with a clear application which often appears to set off the fast exploitation. (Past examples might include linear programming, critical path analysis, and indexed sequential file handlers. On the other hand, relation database methods and word processing hung around for a long time before a lot of people suddenly took an interest.) Consider as possibilities:

Sympathetic user interfaces of several types are being developed. These include 'standard' terminals which will allow a user to communicate with a number of disparate information systems in a single language; interface processors for data centres to allow the use of a common command language by many users; intelligent terminals to make the tasks of searching less tedious by storing protocols and standard inquiries for regular use. References (10,11)

Research is going on in several centres on the use of morphological analysis in framing and improving queries, and in guiding their subsequent processing. This approach consists in essence of using the <u>form</u> and <u>syntax</u> of natural language problem statements to derive suitable search frames, without taking account of the semantic content of the material presented. (12,13).

A number of workers are engaged in studies of automatic subject classification by clustering of numbers of terms occuring in natural language documents. This work has several branches, and has been going on for some years, but now there are real improvements in computing efficiency, which has been a stumbling block, and the promise of results on a practical scale. (14; includes review and bibliography).

The automatic analysis of texts and development of 'inferential' structures has now reached the stage where systems employing these techniques are beginning to offer real promise as analysers of the information implicit in large collections of documents, where each document in itself offers only scraps of data, more or less meaningless in isolation. Applications are envisaged in diagnostic systems; analysis of hospital patient records; automatic scanning and interpretation of broadcast monitoring data. (15,16).

.... and as a promising 'long shot', note the ASK Project. (17). In this study, the basic approach is to assume that a questioner <u>is</u> a questioner because he has an 'anomalous state of knowledge' in which anomalies are to be satisfied by retrieval of information from a database. This apparently oblique attack has resulted in a number of insights into how queries can be formulated, taking into account semantic associations represented by clustering methods, and shows signs of potential performance better than that of conventional 'best match' retrieval systems. Complementary studies on retrieval system performance may prove to be of value for some years to come, as new approaches to searching and classification are tried out. (18)

Finally - and possibly most important of all on a medium timescale - the availability of microcomputer based communications centres with local storage and the facility for calling on remote back up stores when needed, may well break up the present tendency to centralisation of information units, and lead to a

redistribution of document storage to the sites where it is most used. The valuable concept of 'shared investment' in information, which has lead directly or indirectly to such services as the Library of Congress databases, BLAISE, OCLC, DIALOG and many others, will still be retained, but users will be enabled to draw out their sections of interest for local manipulation, and will incur network costs and the occasional messiness of on line working only when they need to refer to exceptional material.

Maidenhead; June 1981

References; The following references have been chosen both for their content and as useful entry points to the rapidly growing literature of the subject. They should all be readily accessible.

1 Ashford, J H and Matkin, D I. The future for free text systems. Computer Bulletin, September 1979

2 Report of a study of the potential users and application areas for free text information storage and retrieval systems in Britain. PROGRAM v14 no1, January 1980

3 Hall, J L. On-line information retrieval sourcebook. Aslib, London. 1978

4 Ashford, J H. Information management packages on minicomputers. J. Information Science, v2 no1 1980

5 British Computer Society Information Retrieval Specialist Group - Third Research Colloquium, held in the University of Aston, March 1981; papers inpress for publication.

6 Whitehead, J B. Word processing and information management: paper presented at Aslib evening meeting 25 February 1981 - to be published in J. of Documentation

7 SIFT Design specifications: September 1980. Government Institution of Organisation and Management, Norway. (Address for inquiries PO Box 8115 Dep., Oslo 1, Norway)

8 STATUS Reference manual. AERE Harwell, 1980.

9 Lundeen, Gerald. The role of microcomputers in libraries. Wilson Library Bulletin, November 1980

10 Williams, P W. The potential of the microprocessor. Aslib Proceedings, v31 no4. April 1979

11 Jamieson, S.H. The economic implementation of experimental retrieval techniques on a very large scale using an intelligent terminal. Proc. Second International Conference on information storage and retrieval (ACM - SIGIR), 1979, Dallas, Texas.

12 McGregor, D R, and Malone, J R. Efficient implementation of retrieval operations making use of recursively enumerated sets. Proc. BCS - Information Retrieval Specialist Group, - Third Research Colloquium, University of Aston, March 1981.

also The FACT database: a system based on inferential methods. Proc. BCS-ACM Conference in Cambridge, June 1980, published as Oddy, R N and others, Information Retrieval Research, Butterworths, London. 1981

13 Antoniadis, G, and Rouault, J. Morphological analysis as a help for automatic indexing. (In (5))

14 Willett, P. Document clustering using an inverted file approach. J. of Information Science, v2. no5. November 1980

15 Schank, R C. and others. Conceptual information retrieval. In Information Retrieval Research, Butterworths, London, 1981 (as 12b)

16 Hirschman, L. Retrieving time information for natural language texts. (as 15)

17 Belkin, N J, and Oddy, R N. Design study for an anomalous state of knowledge based information retrieval system. British Library R & D Report no 5547, Univ. of Aston, 1979.

18 Sparck Jones, K. Automatic indexing. Reprinted - Progress in documentation: some themes and topics. Aslib, London. 1981. Original in J. of Documentation v30 no4. December 1974.

Appendix A Selected list of applications on minicomputer systems.
These examples are drawn from the list of users of STATUS published by AERE Harwell; The author is aware of parallel applications on CAIRS, FACTFINDER, 3RIP and others, but their vendors do not publish convenient lists of their clients!

Location	Hardware	Main applications	Notes
BNF Metals Technology Ltd	Prime	Abstracts database Company Information	
Building Research Establishment	Prime	Library catalogue Library loans	Over 60,000 records
Finsbury Data Services Ltd	DEC	Financial and industrial information service	On line access to subscribers
Police Scientific Dev. Branch	GEC	Document retrieval	(Home Office)
Orion Corp. Ltd	DEC	Product development database	
Ricardo Consulting Engineers	ICL	Technical abstracts Engine database	
Rutherford Laboratory (SRC)	Prime	Engineering drawing specifications	
UKAEA Culham Laboratory	Prime	Fusion experimental data	

(Reproduced by permission, AERE Harwell Laboratories).

Addendum

(September 1981)

Following is a list of some of the companies who have installed BFMIRA CAIRS, their main applications, etc:

COMPANY	APPLICATIONS	HARDWARE IN USE	SCALE OF DB
British Aerospace	IR/LM/CA/PI/ BG/WP	Texas 990/10 running DX10	11 databases: up to 100,000 records
Smith Kline & French (UK & USA)	IR/LM/CA/PI/ BG/PD/WP	PDP 11/70 running RSTS	13 databases: up to 30,000 records
Corn Products Corporation (USA & Belg)	IR/LM/PI	Texas 990/10 running DX10	3 databases: up to 12,000 records
Agricultural Research Council	IR/CA	Texas 990/12 running DX10	1 database: up to 4,000 records
Rio Tinto Zinc Corp	IR/CA	PDP 11/60 running RSTS	1 database: up to 10,000 records
British Petroleum	IR/CA/PI/CT	Texas 990/10 running DX10	2 databases: up to 7,000 records
Manpower Services Commission	IR	Perkin Elmer 3200	1 database: up to 10,000 records
Unilever (Australia)	IR/LM/CA/PI CT/BG	PDP 11/34 running RSX 11M	1 database: up to 50,000 records
Berec Group Ltd	IR	Texas 990/10 running DX10	1 database: up to 5,000 records
Chloride Technical Ltd	IR/LM	Texas 990/10 running DX10	1 database: up to 5,000 records
Glaxo Ltd	IR/LM/PI/BG/ CA	Prime 750	9 databases: up to 100,000 records

KEY:
- IR = Information retrieval
- CA = Current awareness
- CT = Catalogues
- WP = Word processing
- LM = Library management
- PI = Printed indexes
- BG = Bulletins

ONE YEAR OF EURONET DIANE
EXPERIENCE AND EXPECTATIONS

BARRY MAHON, Head Euronet Launch Team
15, avenue de la Faiencerie, Luxembourg.

SUMMARY

The Euronet DIANE network, the information facility instigated and developed by the Commission of the European Communities commenced commercial service on 31st March 1980.

This paper reviews the first years operations. It describes the telecommunications network and in particular the relationship between the European Commission and the consortium of eleven PTT's who provide the technical facilities. The evolution of Euronet into a public packet switched network for all the European Community Member States and also for countries outside the Communities, is also described.

Concerning the information services, collectively referred to as DIANE, the paper describes the expansion that has taken place in files available since commercial opening and analyses the progressive shift in emphasis from purely bibliographic services to a mix of factual data banks and bibliographic data bases.

The ancillary facilities, referral, document ordering and delivery etc. either operational or under development are also described in particular the activities of the European Communities in developing the Common Command Language and the programme of work designed to overcome the Language barrier.

The paper discusses experience of the Commission and the Member States in the development of new information services, the integration of new forms of interactive information services (Viewdata) and the contribution that Euronet DIANE has made towards the evolving "telematique" society.

INTRODUCTION

In order to understand the origins of a project such as Euronet DIANE it is necessary to appreciate in outline the administrative structure of the European Economic Community. The Community has its legal basis in the Treaty of Rome signed by the original six Member States, acceded to by United Kingdom, Denmark and Ireland in 1973 and by Greece in 1981.

Under the Treaty the Commission of the European Communities was set up to provide the day to day executive, administrative and secretarial role; as such, it is an international civil service. The Commission may propose policies and projects to the Member States but may not initiate major Community policies in its own right. Such was the case with the policy basis for Euronet DIANE. In 1971 the Council of Ministers, the representatives of the governments of the Member States, approved a series of resolutions, which had been proposed earlier by the Commission, the purpose of which were to create a network for the collection and dissemination of Scientific and Technical Information (STI) in the European Community.

In these resolutions neither the form of the network nor STI was defined explicitly. However, in the resolutions the Council created a committee to advise the Commission in carrying out the work. This Committee was subsequently called CIDST (Committee for Information and Documentation in Science and Technology) and consists of representatives nominated by the Member States.

This brief historical background is necessary because it is often erroneously suggested that the Commission has had a dictatorial role in the development of Euronet DIANE. The truth is that the Commission's role has been to find a measure of agreement between the Member States in order to carry out the terms of the 1971 resolutions.

EURONET DIANE
one year of operations

The telecommunications network become available for 'pilot' operations on 15 November 1979 and was put into commercial service on 31 March 1980.

The telecommunications facilities are provided under contract by a Consortium of the PTTs (Post and Telecommunications) of the participating countries, which are the 10 countries of the European Economic Communities and Switzerland. A PTT management committee provides the policy and overall supervisory basis while a Network Management Centre, located in London provides the day to day technical management. The Commission of the European Communities through its Directorate General XIII, assisted by CIDST, is responsible for the management of the overall project (in particular relationships with the 'hosts') and provides through the Euronet Launch Team the operations management. The PTTs are responsible for technical maintenance of the network links which consists of 5 switching nodes located in London, Paris, Rome, Frankfurt and Zurich and access facilities in Brussels, Luxembourg (linked to Paris), Dublin, Amsterdam (linked to London) and Copenhagen (linked to Frankfurt). The technology of the network is packet switching using a modified version of the hardware and software of Transpac the French national packet switching data network.

This separation of responsibilities between the PTTs and the CEC was established from the outset because the Communities did not envisage any long term involvement in provision of telecommunications links - its major interest being the provision of reliable and rapid access to stores of information, as specified in the 1971 resolution. The contract between the CEC and the PTT's allows for the 'takeover' of the telecommunications by the PTTs and this process has already commenced and will accelerate over the next 2/3 years. In France, for example, Euronet has never really existed, its functions are provided by Transpac which is linked to Euronet at Paris.

In the near future DATEX-P will provide similar facilities in Germany and PSS in the United Kingdom. These developments will be followed by national data communications facilities in most other countries so that by 1984/85 the links provided to-day by Euronet will be superseded by various interconnections between national facilities in Europe and for access to other areas by links between European countries and USA, Japan, the Far East etc. To the user this change will only be partly apparent, in essence he or she will

use their present national PTT to provide access to a worldwide data communications network similar to the voice network we use to-day.

In the period of slightly over one year that Euronet has been in commercial operation almost 2000 subscribers have signed contracts for use. They have access, at the time of writing, to 26 host computers who support 270 data bases and data banks of information. The evolution of these services in the first operational year illustrates clearly the wealth of resources which Europe contains and justifies the Communities view that the provision of Euronet would act as a catalyst in the creation of a viable information industry for Europe. On 'opening day' there were 16 hosts providing 96 files; by the end of 1980 there were 23 hosts and 137 files and by early 1982 it is envisaged that there will be more than 40 hosts and over 300 files available. The total number of items (bibliographic references or factual records) available to-day is over 60 million of which 50-60% are European in origin.

The evolution has been remarkable not only for the sheer size of the number of items searchable but for the change from bibliographic to factual. From a base of approximately 90:10 bibliographic: factual in early 1980 to-days ratio is approximately 70:30. This change is particularly apparent in the subject area of business and economics and reflects the growing interest of industry in accurate and easily accessible production and financial information for decision making. Of interest also is the geographic distribution of the supplying centres (as expected by end 1981) shown in the table below:

COUNTRY	NUMBER OF HOSTS
BELGIUM	3
FRANCE	11
LUXEMBOURG	2
GERMANY F.R.	6
SWITZERLAND	1
DENMARK	2
ITALY	5
NETHERLANDS	2
UNITED KINGDOM	5

SUPPORT FACILITIES

In addition to the continuous support provided by the Euronet DIANE Launch Team for users, there are a number of other services and user aids available. The biggest and most pervasive of these is the Online News and Information service available free of charge to all users. This facility, provided in six languages, contains a news file in which the hosts and the PTTs can inform customers of operational changes, holiday periods etc; a complete list of the hosts and data bases/data banks available, a list of PTT contact points and details of technical requirements for using the network. The service is continually updated by the Launch Team and the arrangements are such that an urgent news item can be mounted in six languages within one hour of receipt. Another important support facility is the Common Command Language (CCL). It was realised early in the development of Euronet DIANE that, faced with a multiplicity of hosts and consequently with a number of retrieval languages, users would have difficulty in 'learning' the various procedures. The Commission developed in association with a number of hosts a set of commands which would be 'universal' at least in their syntax if not in their detailed effect. These commands can be implemented by a host either as a replacement for existing commands or via a translation device mounted on his machine. For example to locate a keyword and display the number of occurances the CCL uses 'FIND'; to list a set of keywords for selection the CCL uses 'DISPLAY' and to bring up retrieved references the CCL uses 'SHOW'.

Through this method a user can switch from one host using CCL to another and perform searches using the same commands. To-date seven hosts have implemented CCL and a further 7-10 will provide it in the near future.

Also the network has available a file which gives details of compatible terminals for Euronet, the first file of an envisaged all embracing referral facility which will provide a detailed subject guide to the available information files, a list of document supply centres, details of forthcoming training courses and possibly an electronic mail box for users. Other facilities will make available a multilingual data bank of terminology, multilingual thesauri to aid searching and a pilot service for machine translation of retrieved references and abstracts. Another programme of work undertaken by the Commission will, when the necessary policy and legal considerations have been completed provide an electronic document ordering and delivery facility with stores of machine readable documents and high speed digital facsimile for their reproduction. In fact some of the hosts already on the network provide facilities for on-line ordering of retrieved documents, the orders being processed by 'conventional' photocopy supply agencies such as the British Lending Library or the Technische Hochschule at Hannover.

THE FUTURE

Under the Plan of Action 1978-1980 the Commission Launched a 'call for proposals' to support the development of new information products. This 'call' was answered by 267 proposals and obviously not all of these can be funded. To-date 26 projects have received approval and will result in the creation of a number of new information services. It is hoped that in 1982 another such 'call' can be made and thereby a continuous flow of new or restructured information products will become available. In addition the Commission and CIDST are evaluating the export potential for European information services, initially in North America but with a view to later worldwide distribution.

While these activities aimed at widening information supply have been going on the technical developments taking place in ancillary fields have not been ignored. The most interesting of these are Videotex or Viewdata services designed to provide access to information via the telephone and modified TV sets. A project to interface a Euronet host and videotex terminals via Euronet is almost complete and a number of trials in areas such as agricultural information and energy saving information are in course of development.

CONCLUSION

Euronet DIANE was and is a unique concept. The provision of a European wide publicly available communication facility had never been attempted at least not one in which all the facilities would be available at the same time. The organisational, policy and technical problems were significant but the Commision of the European Communities aided by the Member States have achieved, in a relatively short time, the essential aim of providing European industry, commerce, education and service industries with a working tool for better information access. As we move forward into what has been widely predicted as the information age Europe which has progressively grown poorer in its natural resources, which fuelled the industrial revolution, has taken the first steps towards using its considerable intellectual resources to its best advantage.

FORUM DISCUSSION - USER/SUPPLIER DIALOGUE

A summary, prepared by

Rüdiger BERNHARDT — Gesellschaft für Information und Dokumentation mbH, Herriotstrasse 5, 6000 Frankfurt am Main 71, FRG

Gerd TITTLBACH — Fachinformationszentrum Energie, Physik, Mathematik GmbH, Kernforschungzentrum, 7514 Eggenstein-Leopoldshafen 2, FRG

The purpose of the Forum Discussion was to provide the opportunity for an exchange of views between users and suppliers of information services hardware and software with the aim of achieving better cooperation than previously. The discussion took place at the end of the Specialists' Meeting and the principal participants were as follows:

Chairmen:	R Bernhardt, FRG H Sauter, USA
Suppliers:	R G Cypser, IBM, USA D Morgenstern, Siemens AG, FRG
Users:	R Bree, FRG M Day, USA B Mahon, Luxembourg

To start the discussion the two suppliers' representatives gave a review of major hardware and software developments and of future plans which might influence work in the information field. It was evident that over recent years the manufacturers have achieved significant progress in hardware. Computers have become larger and more powerful, and the capacity of the stores has expanded rapidly, while hardware costs have decreased. Once basic technical problems and the legal requirements of telecommunications had been settled, data communication networks were built up and became operational. The reaction of the manufacturers was to focus their efforts increasingly upon management support systems, computer-computer communications, dialogue systems for common use, and information retrieval systems combined with editing and delivery systems. However, nowadays, office automation is obviously occupying an increasingly significant role in the market.

Hitherto, the development of data processing was driven mainly by technological evolution and not by customer requirements. So there is still a lack not only of application software, but also basic software.

With that background the discussion was set up. User requirements on the one hand and the business interests of the suppliers on the other were described from various points of view. It became clear that users often are not able to precisely describe their demands. But it was recognized that many would-be users are limited in their knowledge of what it is reasonable to request, and that manufacturers are often either not aware of the real problems of the appliers or choose to ignore requirements which happen not to coincide with their cost/benefit thinking. We are in an era where technology is leading application. The user virtually is always behind in his use of technology and very far behind the leading edge of technology.

What of the future? The new user generation seems to be more information minded than its predecessor. User groups are being established and a broad information and communication market is growing. Nevertheless, the situation concerning suppliers' support should not be viewed too optimistically. Although the suppliers confirm that, there is a broad lack of support in our application field, because of the pressure of costs and the increasing number of new application fields which continually appear. Suppliers will not be able to develop special hardware and software tools for applications, especially in a restricted market. Therefore, what the manufacturers will offer in future will be more generalized solutions. To balance this trend there is a need to create strong user groups to inform suppliers about specific requirements, indicating their value and relevance. It has to be realised that, in the manufacturers' view, the information services field constitutes only a limited market, much less significant than the office automation area, for example. It follows from this that major developments will normally occur outside the information and documentation area and significant suppliers' support can only be expected when requirements conform to the mainstream requirements. Therefore all the developments in related application fields must be scrutinized as before, to evaluate and pick out ideas and tools which might be transferred and adapted in the best manner possible for our particular application.

Despite this rather pessimistic assessment, there is some hope that over the next ten years good progress can be expected in different fields, such as:

- improved input/output devices, multifont readers, flat colour displays, intelligent displays, microprocessors, small laser printers, intelligent copying machines
- increased storage capacity, optical disc memories
- database management systems by new concepts
- information retrieval systems considering the compilation of formatted and unformatted data
- software improvements in linguistic data processing
- telecommunications by using common networks and in-house communication.

In the second part of the Forum Discussion the audience had the opportunity of putting further questions to Speakers, Panel Members and other participants. To summarize the extended discussions is very difficult but the following points were in the centre of interest:

- studies of user requirements and market research as a basis of further progress in both hardware and software
- cost and budgeting of information
- public relations and marketing of information services
- managing the overwhelming flood of information, redundancy of information, and the access to the full-text by improved information analysis, evaluation and verification methods
- problems of data acquisition, analysis of data and its organization (who will do it, where will it be done; how to reduce and share the cost)
- international cooperation and promotion of information exchange, two-way flow of information
- acquisition problems of scientific and technical libraries in the light of restricted budgets
- copyright problems involved with document delivery from databases.

This second part of the discussion showed an interesting spectrum of questions covering common, every-day, but nevertheless hard problems. However, while discussions usually cannot solve existing problems they do serve to illuminate them by showing different ways - be they good or bad - whereby more effective solutions can be sought.

LIST OF PARTICIPANTS

Mr J C ANDREWS	Chief Librarian, Ministry of Defence, Old War Office Bldg, Room 028, Whitehall, London SW1, England
Dr G M AHRENHOLZ	Head, I and D Dept, Bayer AG, FE-DID, Q 18, 5090 Leverkusen 1, Fed Rep of Germany
Dr J H ASHFORD*	J H Ashford Associates Ltd, 72 Harrow Lane, Maidenhead, Berkshire SL6 7PA, England
Dr H BECHTEL	Metallgesellschaft AG, Postfach 37 24, 6000 Frankfurt, Fed Rep of Germany
Dipl Phys G BELING	Senior Scientist, SHAPE Technical Centre, P O Box 174, 2501 CD The Hague, Netherlands
Mr R BERNHARDT+	Leiter der Hauptabteilung Datenverarbeitung, GID, Herriotstrasse 5, 6000 Frankfurt am Main 71, Fed Rep of Germany
Mr J P BETHELL+	Head, Scientific & Technical Information Dept, SACLANT ASW Research Centre, Viale San Bartolomeo 400, 19026 La Spezia, Italy
Ir J B D H BOLT	Manager,Information Centre, Fokker BV, P O Box 7600, 1117 ZJ Schiphol, Netherlands
Lt Col H BRAUN+	Dokumentationszentrum Bw, Dezernat A, Friedrich-Ebert Allee 34, 5300 Bonn, Fed Rep of Germany
Mr R K H BREE*	An der Herrenwiese 6, 5330 Königswinter 41, Fed Rep of Germany
Mr P BRISSET	Aérospatiale, 12 rue Béranger, 92320 Chatillon, France
Mr A BRUCE+	Head, Defence Research Information Centre, Station Square House, St Mary Cray, Orpington, Kent BR5 3RE, England
Mrs I BUTTEL	Siemens AG, D AP GE 3, Otto-Hahn-Ring 6, 8000 Munich 83, Fed Rep of Germany
Mrs S BUTZ	Siemens AG, Balanstr 73, 8000 Munich 80, Fed Rep of Germany
Mr J P CHILLAG	Senior Scientific Officer, The British Library, Lending Division, Boston Spa, Wetherby, West Yorkshire LS23 7BQ, England
Dipl Phys M CICHOS	Information Officer, Dokumentationszentrum Bw, Friedrich-Ebert Allee 34, 5300 Bonn 1, Fed Rep of Germany
Dr A COCKX+	Director, Centre National de Documentation Scientifique et Technique, Bibliothèque Royale Albert ler, Boulevard de l'Empereur 4, 1000 Bruxelles, Belgium
Dr C von CONSBRUCH	Head of Division, Fachinformationszentrum Energie, Physik, Mathematik GmbH, 7514 Eggenstein-Leopoldshafen, Fed Rep of Germany
Mr J G COYNE+	Manager, Technical Information Center, Dept of Energy, P O Box 62, Oak Ridge, Tennessee 37830, United States of America
Lt Col A CUFFEZ+	Adjoint Informatique Logistique, Etat-Major Force Aérienne (VSL/I), Quartier Reine Elisabeth, Rue d'Evere, 1140 Bruxelles, Belgium
Mr D R CURRY	Room 4/1, MT, Ministry of Defence, London WC2N 5BL, England
Mr R G CYPSER	IBM Coorporation, 44 South Broadway, White Plains, NY 10601, United States of America
Mr M S DAY+	Director, National Technical Information Service, Dept of Commerce, 5285 Port Royal Road, Springfield, VA 22161, United States of America
Mr A DE DIER	Union Minière S A, Service d'Information et de Documentation, Rue de la Chancellerie 1, 1000 Bruxelles, Belgium
Prof M van DIJK	Bureau Marcel van Dijk, Avenue Louise 409, Bte 1, 1050 Bruxelles, Belgium
Mr R DREWS	c/o SOTEC Software-Technik GmbH & Co KG, Nonnendammallee 101, 1000 Berlin 13, Fed Rep of Germany

Reg Dir Dr K H ECKER	Bruderhofstr 30a, 8000 Munich 70, Fed Rep of Germany
Mr P J EDELENBOS	Head, Documentation and Library Division, Ministry of Transport and Public Works, P O Box 20901 2500 EX The Hague, Netherlands
Dr H J EHLERS	Director, Corporate Planning, c/o ERNST KLETT, P O Box 809, 7000 Stuttgart 1, Fed Rep of Germany
Mr C E EHRENFRIED	Technical Director, Reliability Analysis Center, Griffiss Air Force Base, NY 13441, United States of America
Prof Dr Ing T EINSELE	Technische Universitat München, Arcisstr 21, Postfach 20 24 20, 8000 Munich 2, Fed Rep of Germany
Mr G EMBER*	CISTI, National Research Council of Canada, Montreal Road, Bldg M-55, Ottawa, Canada K1A OS2
Mr K A FENT	IABG mbH, Einsteinstrasse 20, 8012 Ottobrunn, Fed Rep of Germany
Col A F FISCHER	HQ SHAPE (PANDA-X), SHAPE, 7010 Belgium
Mr G FRACKENPOHL	Institut für deutsche Sprache, Postfach 54 09, 6800 Mannheim, Fed Rep of Germany
Dr G GAROFALO	c/o Aeritalia S.A.I.p.a., Strada Privata, 80038 Pomigliano-D'Arco Napoli, Italy
Mr J GEORGET	SNIAS, Subdivision Systèmes, BP 96 78130 Les Mureaux, France
Mr G E GLEADALL	Head of Technical Enquiries Section, Ministry of Defence (PE), Defence Research Information Centre, Room 507, Station Square House, St Mary Cray, Orpington, Kent BR5 3RE England
Mr D GOODE+	Chief Librarian, Royal Aircraft Establishment (PE), Ministry of Defence, Farnborough, Hants GU14 6TD, England
Col (retd)Ir E GRUTZMACHER+	Director, Scientific and Technical Documentation Centre for the Armed Forces (TDCK), Nieuwe Frederikkazerne, Van Alkemadelaan 774, 2397 BB The Hague, Netherlands
Ing Gén J A J GUILLEMINET+	Directeur Adjoint du CEDOCAR, 26 Bld Victor, 75996 Paris Armées, France
Mr U HAHN	Universitaet Konstanz, FG Politik/Verwaltungswissenschaft/ Informationsw, Postfach 5560, 7750 Konstanz 1, Fed Rep of Germany
Mr V E HAMPEL*	Integrated Information Systems, L-275, Lawrence Livermore Laboratory, University of California, P O Box 808, Livermore, California 94550, United States of America
Dipl-Ing K O HANSEN+	Dokumentationszentrum Bw, Dezernat A, Friedrich-Ebert Allee 34, 5300 Bonn 1, Fed Rep of Germany
Mr A J M HEESBEEN	Van der Burchlaan 31, Geb 33 K 610, 2509 LS The Hague, Netherlands
Mrs G HILLYER	Air Force Weapons Laboratory, Chief, Technical Library, Kirtland AFB, New Mexico 87117, United States of America
Dr H HUESMANN	Deutsches Theatermuseum, Galeriestrasse 4a/6, 8000 Munich 22, Fed Rep of Germany
Min Rat A JAENDL	Bavarian State Ministry of Agriculture, Food and Forestry, Ludwigstr 2, 8000 Munich, Fed Rep of Germany
Maj C M JORGE+	Direcção do Serviço de Material da FA, Rua da Escola Politécnica 42, 1200 Lisbon, Portugal
Col D KAYA+	Ministry of National Defence, Dept of Research and Development (ARGE), Ankara, Turkey
Mr J E KING	Company Systems Executive, Dowty Fuel Systems Ltd, Arle Court, Cheltenham, Glos GL51 OTP, England
Mr G KIROUAC+	Senior Advisor, Science and Technology for Development, National Research Council of Canada, Room E-302, M-58, Ottawa, Ontario K1A OR6, Canada
Dr H-O KLEIN	Gesellschaft für Mathematik und Datenverarbeitung, Postfach 1240, 5205 St Augustin 1, Fed Rep of Germany
Dr E KOENIG	Planungsstab FIS 2, Villichgasse 17, 5300 Bonn 2, Fed Rep of Germany
Mr E J KOLB	Principal Army Technical Information Officer, US Army Development and Readiness Command, 5001 Eisenhower Avenue, Alexandria,VA 22333, United States of America

Prof Dr D KRALLMANN	Universitaet Essen - GHS, FB 3/Sprach-und Lit. Wiss, Postfach 103 764, 4300 Essen 1, Fed Rep of Germany
Prof Dr G KRESZE	Technische Universitaet München, Org.-Chem. Institut, Lichtenbergstr 4, 8046 Garching, Fed Rep of Germany
Mr H K KROG+	Managing Director, Norwegian Centre for Informatics, Forskningsveien 1, Oslo 3, Norway
Mr K KUBLENZ	Deputy Director, mbp Mathematischer Beratungs- und Programmierungsdienst GmbH, Am Wall 140, 2800 Bremen, Fed Rep of Germany
Dr H K KULLMER	Statistisches Bundesamt, Bibliothek-Dokumentation-Archiv, Postfach 5528, 62 Wiesbaden, Fed Rep of Germany
Ing G M LACHEZE+	Chef Adjoint du Departement Information Documentation, Aérospatiale, BP No 76, 92152 Suresnes Cedex, France
Mrs A LAMVIK*	University Library, Div A, Norwegian Institute of Technology, Høgskoleringen 1, 7034 Trondheim NTH, Norway
Mr B J LEE	Avionic Systems Analyst, British Aerospace, Warton Aerodrome, Nr Preston, Lancs PR4 1AX, England
Mr H J LEEBEEK*	Institut voor Zintuigfysiologie TNO, Postbus 23, 3769 ZG Soesterberg, Netherlands
Mr G LOHR	Akademie des Sanitaets- und Gesundheitswesen der Bundeswehr, ASW Dokumentationsstelle, Neuherbergstrasse 11, Postfach 450161, 8000 Munich 45, Fed Rep of Germany
Dr K LUDWIG	Bundesministerium der Verteidigung, Postfach 1328, 5300 Bonn 1, Fed Rep of Germany
Mr W MACH	Head of Information Department, Abt BT007, MBB GmbH, Postfach 801220, 8000 Munich 80, Fed Rep of Germany
Mr B MAHON*	Equipe de Lancement EURONET-DIANE, 15 Ave de la Faïencerie, Luxembourg, G.D. of Luxembourg
Dipl Volksw G MANTWILL	HWWA - Institut für Wirtschaftsforschung-Hamburg, Neuer Jungfernstieg 21, 2000 Hamburg 36, Fed Rep of Germany
Mr E MASSEY	Technical Information Officer, Technical Library, P.L.C., Warton Aerodrome, Preston, Lancs PR4 1AX, England
Mr M MCLAREN	Management Services, PERME, Waltham Abbey, Essex EN9 1BP, England
Ms P H MEANS	Chief, Office of Technical Information, Defense Nuclear Agency, Washington DC 20305, United States of America
Dr W MEDER *	Gesellschaft für Information und Dokumentation mbH (GID), Sektion für Technik (SfT), Herriotstrasse 5, 6000 Frankfurt (Main) 71, Fed Rep of Germany
Mrs M MEINDL	Dipl.-Librarian, IABG mbH, Einsteinstrasse 20, 8012 Ottobrunn, Fed Rep of Germany
Mr B MELIS	Senior Information Officer, Koninklijke Shell Laboratorium, Dept CiS/di, Postbox 3003, 1003 AA Amsterdam, Netherlands
Dr K-H MEYER-UHLENRIED	Scientific Officer, EURATOM-CCR, I-21020 Ispra, Varese, Italy
Mr H MINDLIN	Program Manager, Metals and Ceramics Information Center, Battelle-Columbus Laboratories, 505 King Avenue, Columbus OH 43201, United States of America
Mrs M-L MISERQUE	Head of Information and Documentation Dept, Solvay & Cie, 310 rue de Ransbeek, 1120 Bruxelles, Belgium
Col G MORELLI+	Direttore, Centro di Documentazione, Tecnico Scientifica della Difesa, Via Clitunno 33, 00198 Rome, Italy
Dr-Ing D MORGENSTERN*	Siemens AG, DV System Technik DB/DC, Hofmannstrasse 51, 8000 München 70, Fed Rep of Germany
Prof Dr H MUENZEL	Fachbereich 8, Anorganische Chemie und Kernchemie, Technische Hochschule Darmstadt, Hochschulstr 4, 6100 Darmstadt, Fed Rep of Germany
Dr E NEUBAUER	Staatsbibliothek, Preussischer Kulturbesitz Berlin, Potzdamerstr 33, 1000 Berlin 30, Fed Rep of Germany
Mrs F OESTERLEIN	IBM Laboratory, Library and Information Centre 7030 Boeblingen, Fed Rep of Germany
Mr H PAJONK	Information Officer, Flugmedizinisches Institut der Luftwaffe, Postfach 172/KFL, 8080 Fürstenfeldbruck, Fed Rep of Germany
Mr H D PAUL	DIMDI, Weishausstr 27, 5000 Koln 41, Fed Rep of Germany

Lt Col A PEDROSO	Direccào do Serviço de Material da FA, Rua da Escola Politécnica 42, 1200 Lisbon, Portugal
Lt Col R V PRIGENT +	CEDOCAR D-4, 26 Bld Victor, 75996 Paris Armées, France
Mrs M DA PIEDADE ROBERTO	Directorate of Energy, Rua da Beneficencia 241, 1600 Lisbon, Portugal
Dr K RODENBERG	Information Officer, Stiftung Wissenschaft und Politik, 8026 Ebenhausen, Fed Rep of Germany
Mr G ROSENAU	DFVLR Inst. für Flugmechanik, Postfach 3267, 3300 Braunschweig, Fed Rep of Germany
I/Lt S SAILER	Documentalist, Zentrum der Bunderswehr für Innere Führung, v Witzlebenstrasse 17, 5400 Koblenz, Fed Rep of Germany
Dr M O SANTOS	Director of SICT/MT, Ministerio do Trabalho, Praca de Londres, no 2-1, 1000 Lisbon, Portugal
Mr K SAUR	KG Saur-Verlag, Poessenbacherstr 2B, 8000 Munich 71, Fed Rep of Germany
Mr H E SAUTER +	Administrator, Defense Technical Information Center, Defense Logistics Agency, Cameron Station, Alexandria, VA 22314, United States of America
Reg Dir R SCHIFFELS	Deutsches Patentamt, Isartalstr 79, 8000 Munich 70, Fed Rep of Germany
Dr W SCHMITZ-ESSER	Director, Information Marketing, Gruner & Jahr AG, Warburgstrasse 50, 2000 Hamburg 36, Fed Rep of Germany
Director K SCHNEIDER	Gesellshaft für Information und Dokumentation, Sektion für Technik, Herriotstrasse 5, Postfach 710370, 6000 Frankfurt/Main 71, Fed Rep of Germany
Dipl-Ing H H SCHUBERT	MTU-München GmbH, Dept ESD, POB 500640, 8000 Munich 50, Fed Rep of Germany
Reg Dir Dr O SCHULZ	Information Specialist, Dokumentationszentrum der Bw, Friedrich-Ebert Allee 34, 5300 Bonn 1, Fed Rep of Germany
Dipl-Pol D SEYDEL	Head of Documentation Department, Stiftung Wissenschaft und Politik, 8026 Ebenhausen, Fed Rep of Germany
Reg Amt Min E-M SIEBERT	Information Officer, Bundesamt für Zivilschutz, Deutschherrenstr 93,5300 Bonn 2, Fed Rep of Germany
Mrs H SKORCZYK	Akademie des Sanitaets- und Gesundheitswesens der Bw, ASW/Dokumentationsstelle, Neuherbergstrasse 11, Postfach 450161, 8000 Munich 45, Fed Rep of Germany
Mr R T SMITH	Director, Non-destructive Testing Information Analysis Center, P O Box 28510, San Antonio, TX 78232, United States of America
Ing W SONTAG	Technical Information Officer, c/o Bundesamt für Wehrtechnik und Beschaffung - AT III 6 - Postfach 7360, 5400 Koblenz, Fed Rep of Germany
Ir W SOEWANDI	Library of the Delft University of Technology, Doelenstraat 101 2611 NS Delft, Netherlands
Mr K J SPENCER	Chief Librarian and Archivist, Insitute of Geological Sciences, Exhibition Road, London SW7 2DE, England
Mr P STERNEMANN	Deutsche Forschungs- und Versuchsanstalt für Luft- und Raumfahrt E.V., Abt Bibliothekswesen, Linder Höhe, 5000 Köln 90, Fed Rep of Germany
Dr B STRECKER	Head of Department, Dokumentation Pflanzenschutz, K 607, Hoechst AG, Postfach 80 03 20, 6000 Frankfurt 80, Fed Rep of Germany
Dr K A STROETMANN	Informationszentrum, Sozialwissenschaften, Lennestrasse 30, 5300 Bonn 1, Fed Rep of Germany
Dipl-Ing G STUMPF	Technische Universitaet Berlin, Universitaetsbibliothek, Fachdokumentationsstelle Luftverkehr, Marchstr 14, 1000 Berlin 10, Fed Rep of Germany
Ir A S T TAN +	National Aerospace Laboratory (NLR), P O Box 90502, 1006 BM Amsterdam, Netherlands
Mr J THUSS	Documentation Consultant, Springer-Verlag, Neuenheimer Landstrasse 28/30, 6900 Heidelberg, Fed Rep of Germany
Dipl-Ing G TITTLBACH +	Fachinformationszentrum Energie, Physik, Mathematik, GmbH, Kernforschungszentrum, 7514 Eggenstein-Leopoldshafen 2, Fed Rep of Germany

Mme L TRIFANOWA-ROMANOWSKA	Chef du Centre de Documentation, Aérospatiale, 12 rue Béranger, 92320 Chatillon, France
Mrs V VALAITIS	Information Scientist, Guidance and Control Information Analysis Center, IIT Research Institute, 10 West 35th Street, Chicago, IL 60616, United States of America
Mr H VAN TOL	Minstry of Transport and Public Works, P O Box 20901, 2500 EX The Hague, Netherlands
Reg Dir M VOEGTEL	Deutsches Patentamt, Zweibrückenstrasse 12, 8000 Munich 2, Fed Rep of Germany
Dr E VORWERK	Gesellschaft für Mathematik und Datenverarbeitung mbH Bonn, Postfach 1240, Schloss Birlinghoven, 5205 St Augustin 1, Fed Rep of Germany
Miss R G D WADSWORTH	Librarian, National Gas Turbine Establishment, Pyestock, Farnborough, Hants GU14 OLS, England
Dr C WEISKE	Head, Chemie-Information und- Dokumentation, Steinplatz 2, 1000 Berlin 12, Fed Rep of Germany
Mr J W WILD	MBB-ZET, Postfach 80 11 09, 8000 Munich 80, Fed Rep of Germany
Dr P E WULLICH	Documentation Officer, Zentrum der Bw für Innere Führung, Ahrbachstrasse 5, 5431 Boden/WW, Fed Rep of Germany
Mr A YANEZ*	Sous-Directeur Adjoint de l'Information Scientifique et Technique, CEDOCAR, 26 Bld Victor, 75996 Paris Armées, France
Mr A SEIDLER	Siemens AG, D AP GE 3, Otto-Hahn-Ring 6, 8000 Munich 83, Fed Rep of Germany
Dr S ZURNATZIS	Head of Documentation, Fraunhofer-Institut für Hydroakustik, Waldparkstr 41, 8012 Ottobrunn, Fed Rep of Germany

* Authors of Papers presented at the Meeting

\+ Members of the Technical Information Panel

REPORT DOCUMENTATION PAGE

1. Recipient's Reference	2. Originator's Reference	3. Further Reference	4. Security Classification of Document
	AGARD-CP-304	ISBN 92-835-0305-8	UNCLASSIFIED

5. Originator	Advisory Group for Aerospace Research and Development North Atlantic Treaty Organization 7 rue Ancelle, 92200 Neuilly sur Seine, France
6. Title	WHAT SHOULD USERS EXPECT FROM INFORMATION STORAGE AND RETRIEVAL SYSTEMS OF THE 1980's?
7. Presented at	the Technical Information Panel Specialists' Meeting held in Munich, Federal Republic of Germany, 9–10 September 1981.

8. Author(s)/Editor(s)	9. Date
Various	December 1981

10. Author's/Editor's Address	11. Pages
Various	122

12. Distribution Statement	This document is distributed in accordance with AGARD policies and regulations, which are outlined on the Outside Back Covers of all AGARD publications.

13. Keywords/Descriptors

Information systems
Information retrieval
Reviews
Forecasting

14. Abstract

These Proceedings concern an AGARD Technical Information Panel Specialists' Meeting held in Munich in September 1981. The Sessions were as follows. Session I: The information storage and retrieval scene in the beginning of the eighties (a retrospect of significant developments in the last decade and a review of the prospects ahead). Session II addressed current factors which will be influential in the eighties, and Session III was concerned with user expectations. The final session of the meeting, which is summarised, was a Forum Discussion in the form of a user/supplier dialogue.

AGARD Conference Proceedings No.304
Advisory Group for Aerospace Research and Development, NATO
WHAT SHOULD USERS EXPECT FROM INFORMATION STORAGE AND RETRIEVAL SYSTEMS OF THE 1980's?
Published December 1981
122 pages

These Proceedings concern an AGARD Technical Information Panel Specialists' Meeting held in Munich in September 1981. The Sessions were as follows. Session I: The information storage and retrieval scene in the beginning of the eighties (a retrospect of significant developments in the last decade and a review of the prospects ahead). Session II addressed current factors

P.T.O.

AGARD-CP-304

Information systems
Information retrieval
Reviews
Forecasting

AGARD Conference Proceedings No.304
Advisory Group for Aerospace Research and Development, NATO
WHAT SHOULD USERS EXPECT FROM INFORMATION STORAGE AND RETRIEVAL SYSTEMS OF THE 1980's?
Published December 1981
122 pages

These Proceedings concern an AGARD Technical Information Panel Specialists' Meeting held in Munich in September 1981. The Sessions were as follows. Session I: The information storage and retrieval scene in the beginning of the eighties (a retrospect of significant developments in the last decade and a review of the prospects ahead). Session II addressed current factors

P.T.O.

AGARD-CP-304

Information systems
Information retrieval
Reviews
Forecasting

AGARD Conference Proceedings No.304
Advisory Group for Aerospace Research and Development, NATO
WHAT SHOULD USERS EXPECT FROM INFORMATION STORAGE AND RETRIEVAL SYSTEMS OF THE 1980's?
Published December 1981
122 pages

These Proceedings concern an AGARD Technical Information Panel Specialists' Meeting held in Munich in September 1981. The Sessions were as follows. Session I: The information storage and retrieval scene in the beginning of the eighties (a retrospect of significant developments in the last decade and a review of the prospects ahead). Session II addressed current factors

P.T.O.

AGARD-CP-304

Information systems
Information retrieval
Reviews
Forecasting

AGARD Conference Proceedings No.304
Advisory Group for Aerospace Research and Development, NATO
WHAT SHOULD USERS EXPECT FROM INFORMATION STORAGE AND RETRIEVAL SYSTEMS OF THE 1980's?
Published December 1981
122 pages

These Proceedings concern an AGARD Technical Information Panel Specialists' Meeting held in Munich in September 1981. The Sessions were as follows. Session I: The information storage and retrieval scene in the beginning of the eighties (a retrospect of significant developments in the last decade and a review of the prospects ahead). Session II addressed current factors

P.T.O.

AGARD-CP-304

Information systems
Information retrieval
Reviews
Forecasting

which will be influential in the eighties, and Session III was concerned with user expectations. The final session of the meeting, which is summarised, was a Forum Discussion in the form of a user/supplier dialogue.

Copies of papers presented at the Technical Information Panel Specialists' Meeting held in Munich, Federal Republic of Germany, 9–10 September 1981.

ISBN 92-835-0305-8

which will be influential in the eighties, and Session III was concerned with user expectations. The final session of the meeting, which is summarised, was a Forum Discussion in the form of a user/supplier dialogue.

Copies of papers presented at the Technical Information Panel Specialists' Meeting held in Munich, Federal Republic of Germany, 9–10 September 1981.

ISBN 92-835-0305-8

which will be influential in the eighties, and Session III was concerned with user expectations. The final session of the meeting, which is summarised, was a Forum Discussion in the form of a user/supplier dialogue.

Copies of papers presented at the Technical Information Panel Specialists' Meeting held in Munich, Federal Republic of Germany, 9–10 September 1981.

ISBN 92-835-0305-8

which will be influential in the eighties, and Session III was concerned with user expectations. The final session of the meeting, which is summarised, was a Forum Discussion in the form of a user/supplier dialogue.

Copies of papers presented at the Technical Information Panel Specialists' Meeting held in Munich, Federal Republic of Germany, 9–10 September 1981.

ISBN 92-835-0305-8

AGARD

NATO OTAN

7 RUE ANCELLE · 92200 NEUILLY-SUR-SEINE

FRANCE

Telephone 745.08.10 · Telex 610176

DISTRIBUTION OF UNCLASSIFIED AGARD PUBLICATIONS

AGARD does NOT hold stocks of AGARD publications at the above address for general distribution. Initial distribution of AGARD publications is made to AGARD Member Nations through the following National Distribution Centres. Further copies are sometimes available from these Centres, but if not may be purchased in Microfiche or Photocopy form from the Purchase Agencies listed below.

NATIONAL DISTRIBUTION CENTRES

BELGIUM
Coordonnateur AGARD – VSL
Etat-Major de la Force Aérienne
Quartier Reine Elisabeth
Rue d'Evere, 1140 Bruxelles

CANADA
Defence Science Information Services
Department of National Defence
Ottawa, Ontario K1A OK2

DENMARK
Danish Defence Research Board
Østerbrogades Kaserne
Copenhagen Ø

FRANCE
O.N.E.R.A. (Direction)
29 Avenue de la Division Leclerc
92320 Châtillon sous Bagneux

GERMANY
Fachinformationszentrum Energie,
Physik, Mathematik GmbH
Kernforschungszentrum
D-7514 Eggenstein-Leopoldshafen 2

GREECE
Hellenic Air Force General Staff
Research and Development Directorate
Holargos, Athens

ICELAND
Director of Aviation
c/o Flugrad
Reykjavik

ITALY
Aeronautica Militare
Ufficio del Delegato Nazionale all'AGARD
3, Piazzale Adenauer
Roma/EUR

LUXEMBOURG
See Belgium

NETHERLANDS
Netherlands Delegation to AGARD
National Aerospace Laboratory, NLR
P.O. Box 126
2600 A.C. Delft

NORWAY
Norwegian Defence Research Establishment
Main Library
P.O. Box 25
N-2007 Kjeller

PORTUGAL
Direcção do Serviço de Material
da Forca Aerea
Rua da Escola Politécnica 42
Lisboa
Attn: AGARD National Delegate

TURKEY
Department of Research and Development (ARGE)
Ministry of National Defence, Ankara

UNITED KINGDOM
Defence Research Information Centre
Station Square House
St. Mary Cray
Orpington, Kent BR5 3RE

UNITED STATES
National Aeronautics and Space Administration (NASA)
Langley Field, Virginia 23365
Attn: Report Distribution and Storage Unit

THE UNITED STATES NATIONAL DISTRIBUTION CENTRE (NASA) DOES NOT HOLD STOCKS OF AGARD PUBLICATIONS, AND APPLICATIONS FOR COPIES SHOULD BE MADE DIRECT TO THE NATIONAL TECHNICAL INFORMATION SERVICE (NTIS) AT THE ADDRESS BELOW.

PURCHASE AGENCIES

Microfiche or Photocopy
National Technical
Information Service (NTIS)
5285 Port Royal Road
Springfield
Virginia 22161, USA

Microfiche
Space Documentation Service
European Space Agency
10, rue Mário Nikis
75015 Paris, France

Microfiche
Technology Reports
Centre (DTI)
Station Square House
St. Mary Cray
Orpington, Kent BR5 3RF
England

Requests for microfiche or photocopies of AGARD documents should include the AGARD serial number, title, author or editor, and publication date. Requests to NTIS should include the NASA accession report number. Full bibliographical references and abstracts of AGARD publications are given in the following journals:

Scientific and Techr al Aerospace Reports (STAR)
published by NASA Scientific and Technical
Information Facility
Post Office Box 8757
Baltimore/Washington International Airport
Maryland 21240, USA

Government Reports Announcements (GRA)
published by the National Technical
Information Services, Springfield
Virginia 22161, USA

Printed by Technical Editing and Reproduction Ltd
Harford House, 7–9 Charlotte St, London W1P 1HD

ISBN 92-835-0305-8

EN

DAT
FILM

4-

DT